

La mecánica de la desinformación

ULISES CORTÉS

Dadme un punto de apoyo y moveré la Tierra
Arquímedes

La pulsión de los seres humanos por transformar su realidad nos ha llevado a realizar obras increíbles de ingeniería, desde la construcción de Stonehenge a los primeros asentamientos en Jericó, Çatal Hüyük, Göbekli Tepe o Uruk, etc. La tradición oral nos ha dejado ecos de la historia de esas magníficas ciudades, la epopeya de Gilgamesh aún resuena en nuestros oídos y en otros textos.

Entonces, los mensajes de los gobernantes, verdaderos o no, estaban inscritos en piedra y fueron pensados para perdurar a través de los siglos. Solo basta pensar en el impacto y en la prolongada permanencia entre nosotros del código de Hammurabi y sus 228 normas. En aquella época, fijar una idea requería de un tiempo y de una capacidad económica y tecnológica nada despreciable. Ya desde entonces el medio era el mensaje.

El cambio a soportes más asequibles, el barro, el papiro, el pergamino, el papel y, desde allí, a los formatos digitales, ha sido vertiginoso, como también lo ha sido la popularización de la enseñanza de la escritura y la lectura. Tal como McLuhan argumentó con razón, el impacto de la imprenta permitió la producción masiva de libros, lo que llevó de forma paulatina a una mayor alfabetización de la población y a la difusión de ideas, así como a un cambio en la forma en que la gente percibía el mundo. Este cambio tecnológico, según él, favoreció la estandarización, la uniformidad cultural, la automatización de la difusión del conocimiento y la investigación científica, todo un fenómeno social y cultural al que denominó la Galaxia Gutenberg¹.



Hoy, la tecnología que soporta las redes sociales, como ya había predicho McLuhan, ha llevado la evolución de las tecnologías de la comunicación a un nuevo extremo: cualquier mensaje, sin importar su veracidad o relevancia, puede propagarse de forma instantánea y alcanzar a millones de personas en cuestión de segundos. Si antes inscribir una idea en piedra era un privilegio de pocos, ahora la inmediatez y el alcance de los medios digitales han popularizado la expresión, al tiempo que han facilitado la posibilidad de influir en la opinión pública. Pero también han abierto las puertas al abuso y la desinformación². El hombre más poderoso del mundo en 2025 es desinformador contumaz. En este contexto, el medio sigue siendo el mensaje, pero el reto actual para el receptor es discernir entre la verdad y la manipulación en un océano de información efímera y viral.

Los humanos persistimos en mudar la realidad, esta vez sin hazañas ingenieriles. No hay que mover grandes piedras, basta con desinformar al vecino y que éste propague un bulo, o una imagen o noticia falsa.

Las mecánicas de la desinformación comprenden un conjunto de técnicas y estrategias diseñadas para manipular la percepción pública y dislocar nuestra concepción del mundo. La desinformación opera en las redes sociales y los medios de comunicación a través, por ejemplo, de usar la negación para desacreditar la información objetiva que resulta perjudicial para los intereses del emisor, utilizando argumentos endebles o ataques personales a fin de restar credibilidad a fuentes legítimas.

También se usa la distorsión que implica modificar o tergiversar hechos y datos, presentando versiones alteradas de la realidad para confundir a la audiencia y justificar acciones propias; o la distracción que funciona desviando la atención del público de temas relevantes o incómodos, introduciendo información falsa o irrelevante que capte el foco mediático y social, desplazando así el foco del debate; otra técnica consiste en provocar la inquietud sembrando dudas, miedo o incertidumbre, alimentando la polarización y la desconfianza en las instituciones, los medios o los actores políticos, para desestabilizar sociedades y dificultar la reacción coordinada frente a la desinformación. Todo esto mediante múltiples canales, formatos y algoritmos para maximizar su impacto y dificultar su detección y refutación.



Los humanos persistimos en mudar la realidad, esta vez sin hazañas ingenieriles. No hay que mover grandes piedras, basta con desinformar al vecino y que éste propague un bulo, o una imagen o noticia falsa.



La irrupción de las tecnologías de Inteligencia Artificial Generativa como ChatGPT, DALL-E, DeepVoice, Midjourney y Synthesia han acelerado y transformado radicalmente la producción y difusión de relatos, multiplicando de manera exponencial la generación de artefactos discursivos que resultan cada vez más indistinguibles al ojo humano y, por tanto, difíciles de diferenciar de la verdad. Las grandes cantidades de información publicadas cada segundo y su acelerada difusión hace que la moderación manual de los contenidos resulte inviable, en particular si se trata de amenazas de alto riesgo.

Un aspecto sórdido en esta cuestión es saber por qué las compañías que producen estas tecnologías no etiquetan sus resultados con una *marca de agua* indeleble que evite la desinformación. Tampoco me queda claro por qué los gobiernos no les obligan a hacerlo. Según Byung-Chul Han, vivimos en una *sociedad de la información* que ha sacrificado la narrativa profunda y reflexiva en favor de una comunicación instantánea, fragmentaria y superficial. La prisa por ser el primero en publicar o comentar no solo acelera la circulación de mensajes, sino que contribuye a una cacofonía digital donde el sentido se diluye y la verdad se hace *difusa*.

En este espacio digital saturado, las voces de los gobiernos no son menores. Bien al contrario, algunos aprovechan los recursos tecnológicos, en especial las herramientas basadas en IA, para manipular narrativas y modelar la opinión pública. El discurso de odio se usa a menudo con la intención de manipular a las masas, especialmente en el ámbito político, y generar desprecio en la población hacia determinados grupos o personas, habitualmente en función de su raza, género, religión u orientación sexual.

La irrupción de las tecnologías de Inteligencia Artificial Generativa como ChatGPT, DALL-E, DeepVoice, Midjourney y Synthesia han acelerado y transformado radicalmente la producción y difusión de relatos, multiplicando de manera exponencial la generación de artefactos discursivos que resultan cada vez más indistinguibles al ojo humano y, por tanto, difíciles de diferenciar de la verdad.

Una exposición reiterada a este tipo de contenido contribuye en la deshumanización de las comunidades implicadas, tal y como se mostró en el genocidio cometido en Myanmar, en 2017, cuando Facebook³ amplificó la difusión de mensajes de incitación a la violencia hacia el pueblo Rohingya, lo que dio como resultado la muerte de más de 10000 civiles. Así, mediante estrategias comunicativas y psicológicas, logran distraer, inquietar o polarizar a la sociedad, profundizando el ruido informativo y debilitando la capacidad colectiva de construir un relato común y veraz. Cada vez la Galaxia Gutenberg está más alejada.

La reciente entrada en vigor del reglamento europeo de IA (AI Act⁴) y su próxima aplicación integral representan, creo yo, un avance decisivo hacia la protección del espacio público digital. Este marco normativo obliga a etiquetar los contenidos generados o manipulados por IA —en particular los denominados *deepfakes*⁵— y exige a las plataformas el desarrollo de herramientas de detección y transparencia, así como la habilitación de canales accesibles para denunciar intentos de desinformación. Estos esfuerzos de regulación representan unos primeros brotes verdes bien fundados para la esperanza.

No podemos olvidar la advertencia que Gabriel García Márquez formuló mucho antes de la existencia de las redes sociales: *Dadme un prejuicio y moveré el mundo*. Esta frase cobra hoy una relevancia especial, pues evidencia cómo los prejuicios y la desinformación pueden propagarse con una velocidad y alcance sin precedentes en el entorno digital. Por ello, aunque el AI Act es un paso crucial, la verdadera protección del espacio público requiere no solo de leyes y tecnología, sino también de una ciudadanía crítica y consciente del poder que tienen quienes poseen los medios en la era de la inteligencia artificial.

ULISES CORTÉS

Catedrático de Inteligencia Artificial de la Universitat Politècnica de Catalunya. Coordinador Científico del grupo High-Performance Artificial Intelligence del Barcelona Supercomputing Center. Miembro del Observatori d'Ètica en Intel·ligència Artificial de Catalunya y del Comitè d'Ètica de la Universitat Politècnica de Catalunya. Es miembro del comité ejecutivo de EurAI. Participante como experto de México en el grupo de trabajo Data Governance de la Alianza Global para la Inteligencia Artificial (GPAI). Doctor Honoris Causa por la Universitat de Girona.



REFERENCIAS

- 1 McLuhan, Marshall. *The Gutenberg galaxy. The Making of Typographical Man*, Toronto (1962). <https://www.ocopy.net/wp-content/uploads/2017/10/mcluhan-marshall-the-gutenberg-galaxy-the-making-of-typographic-man.pdf>
- 2 La desinformación es información falsa o engañosa que se difunde de manera intencionada. Esto se diferencia de la información falsa o engañosa que se crea y difunde por error o sin ánimo de engañar
- 3 <https://www.amnesty.org/es/latest/news/2022/09/myanmar-facebook-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/>
- 4 <https://artificialintelligenceact.eu/>
- 5 Un *deepfake* es un archivo de video, imagen o audio generado o manipulado mediante una aplicación basada en tecnologías propias de la inteligencia artificial, en particular, usando técnicas de aprendizaje profundo (*deep learning*), con el objetivo de crear contenido falso que resulta extremadamente realista y difícil de distinguir del original.

