

De los memes a los 'deepfakes': ¿hacia dónde va la desinformación generada con IA?

En los últimos tiempos, la inteligencia artificial se está convirtiendo en un catalizador para la dopamina que **alimenta las batallas culturales**, desarrollando una esfera pública donde no importa tanto que sea mentira lo que se comparte, sino su **capacidad para alimentar relatos e imaginarios**. Más que una batalla ideológica, está en juego el desafío de diferenciar estructuralmente entre los hechos y las narrativas alternativas a los hechos.

RAÚL MAGALLÓN

El apagón del pasado 28 de abril puso en evidencia cómo la conversación primaria -el cara a cara físico- fue fundamental para reducir la incertidumbre y el miedo social, pero también para mitigar la necesidad de una información actualizada permanentemente.

La desinformación -y sus ramificaciones- circularon, pero con limitaciones y, sobre todo, con una suspensión temporal en torno al contrato de veracidad que implicaba el desconcierto generado por un acontecimiento “histórico”.

Como explicaba Delia Rodríguez, “estamos muy bien equipados contra el rumor tradicional, y le damos a la comunicación oral informal el valor que tiene:

relativo. El rumor digital, incontrolable y más sensible a manipulaciones interesadas, es otra cosa” (Rodríguez, 2025).

Como en el caso de la DANA, la desinformación de carácter ideológico empezó a circular con mayor intensidad a partir del tercer día. En fenómenos disruptivos, y tras el desconcierto en la comprensión de acontecimientos imprevistos, emerge con fuerza la polarización y la instrumentación.

En este escenario, desgraciadamente, la inteligencia artificial (IA) está asumiendo un papel protagonista. Fundéu, que prefiere hablar de “ultrafalso”, señala que la palabra “*deepfake*” “alude a los sistemas informáticos que permiten,

mediante técnicas de inteligencia artificial, desarrollar vídeos manipulados extremadamente realistas, aunque también es frecuente que se aplique a los vídeos así creados. El realismo es tal que puede ser imposible saber que ha sido falseado, lo que sirve, por ejemplo, para propagar contenidos falsos con apariencia de noticias y como 'pornoenganza'. Este es su uso original, pero en ocasiones se utiliza para manipulaciones similares, como en audio"¹.

La estrategia de diferentes actores políticos es utilizar la ficción y los memes como alternativa narrativa a los hechos, el periodismo crítico y la información veraz

En su inicio, los *deepfakes* se utilizaban para el desarrollo de falsos contextos y se servían principalmente de personajes políticos, cantantes y artistas femeninas, para tratar narrativas vinculadas a los ciclos de actualidad. Se podían identifi-

car porque se caracterizaban por la ralentización de movimientos; y, además, destacaban las alteraciones faciales, se reproducía la explotación de estereotipos y generalmente eran compartidas desde cuentas no oficiales (Garriga *et al.*, 2024).

Sin embargo, parte de esto ha cambiado rápidamente y los indicios de imágenes o vídeos creados con IA se basan actualmente en voces robóticas que no modulan la entonación, movimientos de cabeza repetitivos, textura de la piel uniforme, etc. (Maldita, 2025).

En los últimos tiempos, la IA se está convirtiendo en un catalizador para la dopamina que alimenta las batallas culturales, desarrollando una esfera pública donde no importa tanto que sea mentira lo que se comparte, sino su capacidad para alimentar relatos e imaginarios.

Cinco días después del "Gran Apagón" en España, la cuenta oficial de la Casa Blanca publicaba una imagen creada con IA de Donald Trump vestido de papa² y, poco después, compartía otra imagen vestido de maestro Jedi, en alusión a *La Guerra de las Galaxias*³.

Más allá de la ruptura de determina-

1 Fundéu (25 de abril de 2023). "Ultrafalso, alternativa a 'deepfake'". Fundéu. <https://www.fundeu.es/recomendacion/ultrafalso-alternativa-a-deep-fake/>

2 The White House [@WhiteHouse] (3 de mayo de 2025) [Post]. <https://x.com/WhiteHouse/status/1918502592335724809>

3 The White House [@WhiteHouse] (4 de mayo de 2025). Happy May the 4th to all, including the Radical Left Lunatics who are fighting so hard to to bring Sith Lords, Murderers, Drug Lords, Dangerous Prisoners, & well known MS-13 Gang Members, back into our Galaxy. You're not the Rebellion - you're the Empire. May the 4th be with you [Post]. <https://x.com/WhiteHouse/status/1919053040734072844>

das reglas de comunicación política, la estrategia que se ha ido consolidando por diferentes actores políticos en los últimos años es utilizar la ficción, los memes y las referencias a la cultura popular como alternativa narrativa a los hechos, el periodismo crítico y la información veraz.

El desarrollo de los *deepfakes* plantea más dudas que certezas en relación con la alfabetización mediática, la legislación y los mecanismos de identificación automatizados

De esta forma, y en poco menos de dos años, hemos visto cómo debatíamos sobre su uso experimental por parte de los medios de comunicación a la existencia de páginas como *MrDeepFakes* con más de 650.000 miembros (*Bellingcat*, 2025).

Su facilidad de acceso y de uso se ha acelerado paralelamente al desarrollo de comunidades que lo explotan ideológica o económicamente. En 2024, un análisis de Channel 4 News de los cinco sitios web de *deepfakes* más visitados reveló que en ellos aparecían casi 4.000 personas famosas.

Asimismo, es creciente el uso de *deepfakes* para desarrollar narrativas y estrategias de comunicación corporativa o

para el intento de timos y estafas a través de la suplantación de voz o imagen en tiempo real, por lo que se trata de un fenómeno en el que hay que seguir explorando sus consecuencias en diferentes campos y desarrollar campañas de concienciación pública.

Como ocurrió hace años con internet, estamos en un momento en el que hay que distinguir entre el acceso a la IA y su uso. La diferencia es que esta vez el acceso y el uso se han acelerado mucho más de lo previsible hasta hace muy poco.

Al respecto, el desarrollo de los *deepfakes* plantea más dudas que certezas en relación con cuestiones como la alfabetización mediática, la legislación y los mecanismos de identificación automatizados. Además, el uso ilícito relacionado con desnudos y pornografía por parte de las generaciones más jóvenes incorpora debates que van más allá de su propia naturaleza tecnológica.

Como ya ocurrió con internet, la normalización de su uso -sobre todo, por parte de los más jóvenes- hace que su consumo se produzca mucho antes de que la educación formal debata y explique sus limitaciones y disfunciones sociales.

¿Qué medidas han tomado las redes sociales y plataformas contra la desinformación creada por los usuarios con IA?

Para responder a esta pregunta primero es importante recordar que la propia filo-

sofía de las redes sociales y plataformas se basa en el uso de sus propios sistemas de inteligencia artificial para recomendar contenidos personalizados, pero también para intentar identificar contenidos que vulneren sus propias reglas.

En teoría, X (antes Twitter), Meta, TikTok y YouTube (Google) indican en sus políticas que los usuarios deben etiquetar contenido generado con IA. Sin embargo, sigue sin haber una concienciación de los usuarios de que el proceso automatizado de identificación no funciona suficientemente bien.

Además, la moderación humana de estas plataformas está muy limitada por la gran cantidad de contenidos que se publican, y la incentivación de las notas de la comunidad por parte de las redes sociales no está teniendo el suficiente impacto a la hora de identificar y mitigar el alcance de contenidos potencialmente perjudiciales o engañosos.

En cualquier caso, la breve experiencia nos indica que son necesarias etiquetas claras que indiquen cuándo una imagen ha sido creada con inteligencia artificial.

En la actualidad, entre las medidas que se debaten está la desmonetización de cuentas que utilicen la IA sin identificar para generar ingresos, la prohibición de *deepfakes* políticos o pornográficos en normativas legislativas específicas y

el desarrollo de marcas de agua (in)visibles para poder ser identificadas rápidamente por sistemas automatizados.

Sigue siendo fundamental el papel de las organizaciones de *fact-checking* en la identificación y reducción del impacto de los *deepfakes*

También hay iniciativas como la C2PA (Coalition for Content Provenance and Authenticity) que trabajan en la procedencia de los datos básicos y fiables sobre el origen de un contenido digital (imagen, vídeo, audio, documento, etc.)⁴. Puede incluir información como quién ha creado el archivo, cómo, cuándo y dónde se creó o si se editó.

Así, en teoría, el autor del contenido siempre puede tener el control sobre si se incluyen los datos de procedencia, así como sobre qué datos se incluyen.

La tecnología más allá del mito de progreso es una promesa de sociedad

En esta carrera que identifica progreso y promesa de una nueva sociedad, hay otras cuestiones que parecen ir a paso de cangrejo. La experiencia nos está demostrando que sigue siendo fundamental el papel de las organizaciones

4 Coalition for Content Provenance and Authenticity (s.f.). "Frequently Asked Questions". C2PA. <https://c2pa.org/faq/>

de *fact-checking* en la identificación y reducción del impacto de los *deepfakes*, la limitación contextual de las notas de la comunidad y, asimismo, que las plataformas podrían tener mayor capacidad de respuesta ante los contenidos dudosos o engañosos, paralizando temporalmente la viralización de este tipo de contenidos hasta que sean etiquetados correctamente.

En línea con lo demostrado durante las últimas décadas, las plataformas siguen necesitando mayores estándares de transparencia, rendición de cuentas y exhaustividad sobre esta cuestión en sus informes periódicos, así como el desarrollo de iniciativas de alfabetización mediática y la incorporación de herramientas de identificación disponibles para los propios usuarios.

Desde el punto de vista legislativo, cabe recordar que en el ámbito europeo está en vigor la Ley de Servicios Digitales (DSA) -aunque su cumplimiento sigue siendo muy limitado en esta cuestión- a la espera de que el Código de Conducta sobre Desinformación de carácter voluntario entre en vigor el 1 de julio de este año y permita una mejor rendición de cuentas⁵.

Entre las posibles debilidades del es-

cenario más próximo sigue estando el acceso a datos por parte de los investigadores, mejorar sus sistemas de transparencia -incluidos los relacionados con el uso de inteligencia artificial- y potenciar los mecanismos de coordinación y corrección público-privados.

En el ámbito estatal queda por ver cómo se puede profundizar en el derecho de rectificación en contenidos digitales, qué medidas son necesarias desde el punto de vista de los procesos electorales para mitigar su impacto en la opinión pública y el papel de los órganos independientes creados para monitorizar e identificar este tipo de riesgos sistémicos.

En palabras de Rafael Rubio (2025), “el diseño de la tecnología es también el diseño de la sociedad”. En juego, más que una batalla ideológica, está el desafío de diferenciar estructuralmente entre los hechos y las narrativas alternativas a los hechos, pero también el reto de limitar las posibilidades de alimentar esas narrativas.

Si queremos seguir defendiendo sociedades abiertas, tendremos que empezar a demandar también tecnologías más abiertas y transparentes que permitan una rendición de cuentas más ágil. ■

5 European Commission (13 de febrero de 2025). *The Code of Conduct on Disinformation*. <https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation>

REFERENCIAS BIBLIOGRÁFICAS

- Garriga, M., Ruiz-Incertis, R., y Magallón-Rosa, R. (2024). “Artificial intelligence, disinformation and media literacy proposals around deepfakes”. *Observatorio (OBS*)*, 18(5). <https://doi.org/10.15847/obsOBS18520242445>
- *Bellingcat* (7 de mayo de 2025). “Unmasking MrDeepFakes: Canadian Pharmacist Linked to World’s Most Notorious Deepfake Porn Site”. *Bellingcat*. <https://www.bellingcat.com/news/2025/05/07/canadian-pharmacist-linked-to-worlds-most-notorious-deepfake-porn-site/>
- *Maldita.es* (2025). “Cómo perfiles de TikTok usan IA generativa para desinformar sobre supuestos casos de corrupción destapados en EE. UU. que afectarían a Sánchez y su Gobierno”. *Maldita.es*. <https://maldita.es/malditobulo/20250311/videos-tiktok-sanchez-corrupcion-estados-unidos-vox/>
- Rodríguez, D. (2025). “Apagón desinformativo: ni luz ni bulos”. *El País*. <https://elpais.com/opinion/2025-05-01/apagon-desinformativo-ni-luz-ni-bulos.html>
- Rubio Núñez, R. (2025). “El uso de la inteligencia artificial en las campañas electorales y sus efectos democráticos”. *Revista de Derecho Político* (122), 65–102. <https://doi.org/10.5944/rdp.122.2025.44742>