# Contenido generado por inteligencia artificial: oportunidades y amenazas

## *AI-generated content: opportunities and threats*

**Jorge Franganillo**
https://orcid.org/0000-0003-4128-6546

*Universidad de Barcelona*
*Facultad de Información y Medios Audiovisuales*
*Centro de Investigación en Información, Comunicación y Cultura (CRICC)*
franganillo@ub.edu

---

Esta nota *ThinkEPI* es un texto provisional para el *Anuario ThinkEPI* que se distribuye con la finalidad de abrir su contenido al debate como sistema de revisión abierta. Esperamos comentarios y aportaciones, que a buen seguro enriquecerán el texto y generarán respuestas por parte del propio autor y de otros lectores. Se ruega que no se reproduzca en blogs u otros medios ya que se trata de una versión provisional que podrá ser modificada por el autor y los editores para su publicación definitiva en acceso abierto en el *Anuario ThinkEPI 2022*, una iniciativa de Ediciones Profesionales de la Información S.L.

thinkepi@gmail.com
https://thinkepi.profesionaldelainformacion.com

**1 septiembre 2022**

---

**Resumen:** En los últimos años se ha visto un crecimiento exponencial de los desarrollos orientados a la creación de contenido textual, gráfico, sonoro y audiovisual mediante inteligencia artificial. Son logros tecnológicos extraordinarios que ofrecen grandes oportunidades potenciales, pero a la vez quedan expuestos a ciertos usos cuestionables que son, en sí mismos, amenazas. Este artículo examina iniciativas recientes dirigidas a la simulación de la escritura humana, la creación de vídeos *deepfake*, la clonación de voz y la generación de imágenes a partir de indicaciones textuales. De estos desarrollos se comentan los beneficios potenciales y los principales peligros derivados de un uso incorrecto.

**Palabras clave:** Creación de contenido; Inteligencia artificial; Aprendizaje profundo; Entretenimiento; Desinformación; Propiedad intelectual; Ética de la tecnología.

**Abstract:** In recent years there has been an exponential growth in developments aimed at creating textual, graphic, aural and audiovisual

content through artificial intelligence. These are extraordinary technological achievements that offer great potential opportunities, but at the same time they are exposed to certain questionable uses that constitute real threats. This paper examines recent initiatives aimed at simulating human writing, creating deepfake videos, voice cloning, and generating images from text prompts. Then the potential benefits of these developments as well as the main dangers derived from their misuse are discussed.

**Keywords:** Content creation; Artificial intelligence; Deep learning; Entertainment; Disinformation; Intellectual property; Ethics of technology.

## 1. Introducción

La ciencia ficción especula sobre el futuro de la tecnología, sobre un futuro altamente tecnológico, sobre un futuro dominado por la tecnología. Así, la literatura, el cine y las series de televisión cautivan al gran público con hipótesis sobre una inteligencia artificial (IA) convertida en un elemento de uso cotidiano, tal vez controlándolo todo. Estas narrativas proyectan ideales de futuro que presentan a la IA como un logro tecnológico con grandes beneficios potenciales, pero también como una amenaza que podría rebelarse contra sus creadores y contra el resto de la humanidad.

Más allá de la ciencia ficción está la realidad, y la IA es una realidad con una presencia destacada en numerosos ámbitos. Por su capacidad para resolver problemas complejos simulando el pensamiento humano, es hoy una parte esencial de la vida cotidiana. Ofrece incontables aplicaciones prácticas en ámbitos tan diversos como la sanidad, las finanzas, la meteorología o el transporte, entre otros (**Boden**, 2018; **Bhargava**; **Sharma**, 2022). En estas áreas, esta tecnología puede desempeñar, con una eficiencia cada vez más notable, actividades que normalmente requerirían capacidades tan humanas como la comprensión, el aprendizaje, el razonamiento y la toma de decisiones.

Una de las aplicaciones de la IA que más debate está suscitando es la destinada a la creación de contenido. Las herramientas de aprendizaje automático han abierto un universo de posibilidades para la producción automática de textos, imágenes, sonidos, música y vídeos a partir de los datos y las indicaciones que se les proporcione. Para ello se suelen emplear redes neuronales, esto es, sistemas informáticos formados por nodos interconectados que actúan como neuronas humanas y que se entrenan mediante procesos de aprendizaje automático (**Campesato**, 2020). Esos procesos incluyen el aprendizaje profundo, que es una arquitectura que imita la forma en que los humanos adquieren ciertas habilidades, tales como el reconocimiento de formas o la predicción de palabras.

En los últimos años, la IA generativa ha acelerado su crecimiento en capacidad tecnológica y está alcanzando un grado de sofisticación hasta hace poco impensable. Pero cabe tener en cuenta que la seguridad que puede ofrecer una herramienta, como bien se sabe, depende del uso que se le dé. Y dado que la creatividad humana es ilimitada, la misma tecnología capaz de resolver viejos problemas puede traer problemas nuevos si no se utiliza de forma adecuada. De ahí que en los últimos años hayan surgido iniciativas con impacto social, tales como el proyecto

*OpenAI* o el observatorio *OdiseIA*, que apuestan por un uso responsable y ético de la IA en beneficio de todo el mundo.

Los recientes avances en IA dibujan, pues, un panorama complejo del que emergen indudables oportunidades, pero también limitaciones y amenazas. Dada la magnitud de los progresos orientados a la producción de contenidos, resulta oportuno examinar estos nuevos desarrollos y los efectos que pueden tener en el público receptor.

## 2. Suplantar la escritura humana

Una de las aplicaciones más prometedoras del aprendizaje automático es la producción de textos que simulan la redacción humana. Con este propósito se han desarrollado varios modelos de lenguaje. El más potente hasta ahora es el denominado *Generative Pre-trained Transformer 3*, más conocido por sus siglas: *GPT-3*. Creado por el laboratorio *OpenAI*, este modelo está diseñado para producir textos que imitan de forma convincente la escritura humana, es decir, la redacción de un texto pensado y escrito por una persona. Tal es la calidad de los textos que así se generan que resulta difícil distinguirlos de aquellos escritos por personas (**Walsh**, 2022).

*GPT-3* no solo produce textos, sino que también es capaz de resumirlos y traducirlos basándose en el estudio del contexto, lo que confiere a este modelo un enorme potencial y un buen número de aplicaciones prácticas. Ya son numerosas las empresas que han lanzado alguna herramienta apoyada en *GPT-3* para que quien quiera pueda elaborar estos textos. Existen aplicaciones específicas en el ámbito del marketing, la literatura y el periodismo, y se está explorando el potencial de este modelo en la escritura científica (**Osmanovic-Thunström**, 2022).

En el terreno de la redacción web, el marketing de contenidos y el *copywriting* han surgido servicios de redacción y corrección automática capaces de generar todo tipo de contenido digital a partir de unas breves y sencillas indicaciones. Aplicaciones web como *Anyword*, *Copy.ai*, *Copymatic*, *Copysmith*, *Jasper*, *Peppertype*, *StoryLab.ai*, *WordAI*, *Wordtune* o *Writesonic* arrojan propuestas rápidas de titulares, lemas comerciales, tuits, artículos de blog, anuncios o descripciones de producto, entre otros textos. Cabe advertir de que tales propuestas no siempre suenan naturales, por lo que conviene revisarlas, elegir los mejores fragmentos de cada una y pulir el conjunto antes de su publicación.

Quienes producen literatura de ficción pueden vencer el temido «bloqueo del escritor» (o «síndrome de la página en blanco») si se ayudan de *Sudowrite*, otro programa de escritura automática basado en *GPT-3*. Basta con pegar en su interfaz un fragmento de texto y la herramienta puede reescribirlo con tono amenazador, o con mayor conflicto interno, o con un giro en la trama, o puede generar descripciones o metáforas en diversos sentidos (**Dzieza**, 2022).

En general, la IA parece dar buen resultado en la creación de textos breves como, por ejemplo, un artículo de opinión para la prensa, tal como el que publicó *The Guardian* (figura 1), o bien algo de poesía, como así lo demostró el proyecto *BNElab*, de reutilización de la Biblioteca Nacional de España (**Nogales**, 2022). Pero todavía falla en construcciones más prolongadas, más complejas, más largas, en las que parece incapaz de seguir un hilo narrativo coherente. Queda aún lejos, pues, el momento

en el que una IA pueda escribir toda una novela o el guion de un largometraje (**Du Sautoy**, 2020).


Figura 1. Artículo escrito por el modelo *GPT-3* con la misión de convencer al público lector de que los humanos no deben temer a la IA. Fuente: https://theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

### 3. Suplantar la realidad: vídeos *deepfake*

En el ámbito audiovisual, el término *deepfake* alude a la técnica de IA capaz de generar vídeos manipulados con un aspecto tan realista que hace difícil notar que han sido falseados. Para ello se emplea una clase de algoritmos conocida como *redes generativas antagónicas*, que a través del aprendizaje profundo pueden generar, sin intervención humana, imágenes o voces que parecen auténticas.

Para referirse a ese contenido hipertrucado, que no es real pero que lo parece gracias a una manipulación extrema, se ha propuesto para el término *deepfake* la traducción *ultrafalso*, que define con acierto a ese tipo de imágenes que hacen creer que una persona dice o hace algo que no es real. El significado del término sin duda denota que las imágenes se pueden falsear en un grado tan extremo que cuesta mucho

verles el truco. El engaño, en efecto, es tan sofisticado que resulta prácticamente indetectable.

La recreación de personas con resultados realistas llegó de la mano del cine, con el uso de la técnica de efectos especiales conocida como CGI (por las siglas en inglés de *imágenes generadas por ordenador*) para recrear a intérpretes fallecidos durante el rodaje, como ocurrió, por ejemplo, con Oliver Reed en *Gladiator* (2000) y con Paul Walker en *Fast & Furious 7* (2015). En estos casos se filmó a un doble y luego el rostro se le reemplazó por medios digitales con el rostro del actor desaparecido.

Otro antecedente notorio es *Forrest Gump* (1994), cuyo logro técnico condujo a que se bautizase como «efecto Forrest Gump» la práctica popularizada por esta película, que consiste en insertar a una persona o un objeto actual en imágenes históricas (**De-la-Cuadra-de-Colmenares; López-de-Solís; Nuño-Moral**, 2014), tal como han hecho después marcas como Peugeot, Mercedes Benz o Virgin Trains en sus anuncios publicitarios (figura 2). Pero estas técnicas tienen el inconveniente de resultar caras, puesto que requieren contar con un personal experto que además debe emplear bastante tiempo.


Figura 2. El anuncio publicitario del Peugeot 406 (1999) es un buen ejemplo del «efecto Forrest Gump».

### 3.1. Tecnología *deepfake* al servicio del entretenimiento y la publicidad

La IA ha supuesto tal avance y se ha vuelto tan accesible que los logros conseguidos con la tecnología CGI nos parecen hoy anticuados. La tecnología *deepfake* ha cambiado las reglas del juego y ha facilitado la recreación de ciertas personas en las industrias creativas. De contenido ultrafalso se pueden mencionar aplicaciones destacables, como, por ejemplo, la producción de material para el cine o la publicidad.

La tecnología *deepfake* ha permitido rejuvenecer a personajes de películas con fines narrativos. Lo ejemplifica el tratamiento realizado con Robert De Niro en escenas de *El irlandés*, Will Smith en escenas de *Géminis*, y Carrie Fisher en la saga de *La guerra de las galaxias*. También ha «resucitado» a celebridades fallecidas, como Lola Flores,

en la campaña «Con mucho acento» de la cervecera Cruzcampo, o Salvador Dalí, en una campaña del Museo Dalí de Florida. Y permite incluso suplantar en tiempo real a personalidades de la esfera pública con fines satíricos, como hace el programa *El Intermedio* (La Sexta) en las secciones «Entrevista por la cara» y «Aznarito y Felipón», donde el presentador Gran Wyoming y el colaborador Dani Mateo encarnan a figuras políticas reconocidas (figura 3).

Esa misma aplicación de la IA ha estimulado, en el ámbito de Internet, la creación y difusión de vídeos divertidos en los que el rostro de carismáticas estrellas, como Nicolas Cage o Steve Buscemi, suplantan a intérpretes, incluso femeninas, de cualquier película. Y también ha dado una enorme popularidad al canal *DeepTomCruise* (@deeptomcruise) de *TikTok*, en el que un imitador sorprende a miles de personas suplantando al actor Tom Cruise en clave de humor.


Figura 3. Gran Wyoming y Dani Mateo suplantan a José María Aznar y Felipe González para analizar la actualidad al estilo de Faemino y Cansado.

### 3.2. La amenaza *deepfake*

La capacidad de falsificar imágenes y sonidos no es nueva. Pero la tecnología que hoy conocemos con el nombre genérico de *deepfake* o, de forma más precisa, *falsificaciones sintetizadas por IA*, delega en los ordenadores el tedioso trabajo que antes se debía hacer a mano. De este modo, en lugar de sentar un técnico ante un ordenador para que retoque manualmente cada fotograma de un vídeo, se entrena a una IA para que haga ese proceso de forma autónoma y automática. Basta darle un conjunto de imágenes y una cara, y el sistema aprende a superponer la cara en las imágenes.

El contenido ultrafalso ha alcanzado tal grado de sofisticación que lo ficticio ya apenas se distingue de lo real. La manipulación resulta difícil de detectar y esta cualidad —la verosimilitud— resulta efectiva en el entretenimiento y en la publicidad. Pero al mismo tiempo supone una amenaza y puede ser destructiva. Es tan importante la capacidad de la tecnología *deepfake* para engañar a cualquier observador que esta tecnología puede llegar a usarse con fines maliciosos, como crear bulos,

falsear noticias, perpetrar estafas y lanzar ataques contra el honor o la reputación de una persona, una institución, un gobierno, etc.

Un detalle que solía delatar a estos contenidos falsificados era el «efecto del valle inquietante», llamado así por la sensación perturbadora que provoca la mirada vacía de una persona artificial. Pero hoy las imágenes son cada vez más convincentes y parecen transportar al observador, desde ese valle inquietante, donde era fácil descubrir la trampa, a un lugar más profundo: un mundo donde el engaño se hace más difícil de descubrir.

El realismo de esas recreaciones ha llegado a un nivel tan asombroso que los rostros producidos por IA ya no solo son indistinguibles de las caras reales, sino que además tienden a generar más confianza (**Nightingale**; **Farid**, 2022). Con lo cual, estos algoritmos, que representan todo un logro técnico, son al mismo tiempo una artillería perversa para toda clase de usos maliciosos. Esas manipulaciones, de tan creíbles, pueden usarse también para el fraude financiero, para difamar y obtener rédito en campañas políticas, para generar imágenes íntimas falsas con miras a un chantaje y, seguramente, para nuevas e insospechadas formas de fraude y abuso. Para más desasosiego, la tecnología destinada a crear falsificaciones sintetizadas es cada vez más accesible, de modo que cualquiera puede crear hoy contenidos artificiales sin necesidad de conocimientos especializados de CGI o de retoque digital.

**4. Suplantar la voz humana**

La síntesis del habla es, en sí misma, una tecnología relativamente simple. Desde hace tiempo desempeña un papel vital como tecnología de apoyo mediante sencillos convertidores de texto a voz hablada que se han usado de forma generalizada en una variedad de escenarios: asistentes de voz, lectores de pantalla para personas con discapacidad visual, navegación GPS, aprendizaje de idiomas y atención telefónica automatizada, por citar algunos ejemplos (**Franganillo**, 2022), lo que demuestra que la tecnología no necesita ser sofisticada para aportar soluciones útiles.

Ahora, últimamente, la IA ha permitido explorar nuevos caminos, uno de los cuales es la posibilidad de clonar voces conocidas para usarlas, siempre de forma ética y bajo licencia, en la locución de productos audiovisuales y de entornos inmersivos. En esta fórmula es pionera la empresa *Veritone*, que a través de su servicio *Veritone Voice* permite nutrir cualquier proyecto sonoro, multimedia o audiovisual con voces de atletas, celebridades, actores, actrices u otras personas influyentes. La ventaja de basarse en una IA es que acelera procesos y recorta gastos. Una vez que se han sintetizado las voces, la personalidad que presta su voz no ha de acudir a propósito a un estudio de grabación, según su disponibilidad, y dedicar unas horas a grabar una locución, sino que cobra unos *royalties* por la voz sintetizada que genera un algoritmo en tiempo real.

Esta aplicación ilustra, pues, una de las funciones de la IA: resolver tareas expertas sin el coste que supone encargar tales tareas a una persona experta en un área de especialización, lo que requeriría pagarle unos honorarios elevados y asumir una disponibilidad limitada. Estas son, de hecho, las principales ventajas de los denominados *sistemas*

*expertos*: actúan como un humano, pero sin las lógicas exigencias de un ser humano.

Paralelamente, existen proyectos basados en IA orientados a preservar la voz de personas con discapacidad en el habla, afectadas por la enfermedad de la motoneurona, por una distrofia muscular, por un cáncer de garganta, etc. (**Dale**, 2022). En este terreno han hecho desarrollos destacables empresas tecnológicas como *Dell Technologies*, *Google*, *Intel* y *Sonantic*.

## 5. Suplantar la creación gráfica: imágenes a partir de texto

Una red neuronal se puede entrenar con un gran corpus de texto para generar imágenes de alta resolución como respuesta a simples indicaciones verbales. Así quedó demostrado en enero de 2021, cuando la empresa *OpenAI* desarrolló *DALL·E*, el primer modelo de IA capaz de generar imágenes originales a partir de una descripción textual. Tal fue su éxito que *OpenAI* lanzó en abril de 2022 una versión actualizada: *DALL·E 2*, capaz de producir imágenes más realistas y con mayor resolución, aunque de acceso todavía restringido, mediante invitación.

La invitación me llegaría un mes después de pedirla, y la experiencia resultante fue tan sorprendente y prometedora que vale la pena dejar aquí una constancia de ella. Para experimentar el nivel de capacidad asociativa entre una entrada textual y un resultado visual, al programa le indiqué que crease imágenes a partir de esta sentencia: «Robotic hands typing on a computer keyboard». Tras unos segundos de espera, la IA ya había creado cuatro imágenes sugerentes y de buena calidad (figura 4). Se trata de imágenes libres de derechos de autor y entregadas en un formato (PNG) y una medida (1024×1024 píxeles) aptos para publicar.

Figura 4. Una de las imágenes obtenidas con *DALL·E 2*, que demuestra el potencial de la IA aplicada a la creación de imágenes a partir de una sencilla línea de texto.

La síntesis de imágenes a partir de texto ha ganado una popularidad enorme, en buena medida gracias a *DALL·E mini*, una alternativa gratuita y abierta al público, rebautizada hoy como *Craiyon* (https://craiyon.com). Esta herramienta, creada por Boris Dayma, es capaz de dar forma a cualquier idea expresada en lenguaje natural. Sea cual sea la sugerencia que se describa con palabras, y por extravagante que pueda sonar, *Craiyon* la traduce a nueve imágenes de baja resolución. Desde que se puso a disposición pública, en junio de 2022, este *software* de código abierto ha recibido mucha atención, en particular por los sorprendentes resultados que arroja cuanto más fantasiosas y precisas son las indicaciones suministradas. La comunidad *r/weirddalle* de *Reddit* recoge las creaciones más delirantes.

También este mismo año, *Google* ha lanzado su proyecto de IA con el que aspira a competir con *DALL·E*. Se trata de *Imagen*, un modelo que destaca por su alto grado de realismo y por el profundo nivel de comprensión del lenguaje. Sin embargo, todos estos sistemas presentan una limitación: al basarse solo en indicaciones textuales, resulta difícil predecir el aspecto que tendrá el resultado. La solución a este problema la ha presentado la empresa *Meta* con un desarrollo paralelo, *Make-A-Scene*, cuyo rasgo diferencial es que permite complementar las indicaciones textuales con bocetos gráficos, que ayudan a precisar la composición final de la imagen.

Estos modelos con capacidad fotorrealista suponen un impulso para la creatividad humana, pero también están expuestos a usos maliciosos. *OpenAI*, *Google* y *Meta* son conscientes de esta realidad. Por esta razón, trabajan de manera cerrada en sus respectivas tecnologías y han decidido no publicar fragmentos de código o demostraciones públicas de sus desarrollos.

Mientras tanto, ya comienzan a verse usos editoriales de este tipo de imágenes, que resultan peculiares, por cierto, por sus tintes surrealistas. La revista británica *The Economist* utilizó el programa *Midjourney* para ilustrar la portada de su número de junio de 2022, que incluía un informe monográfico sobre las promesas y los peligros de las tecnologías de IA.

La progresiva aparición de nuevas herramientas, como el modelo de código abierto *Stable Diffusion*, y de nuevas características, como la función *Outpainting* de *DALL·E 2*, capaz de ampliar la superficie de una imagen más allá de sus límites originales, hace razonable prever que el uso y las aplicaciones de esta tecnología irán en aumento.

## 6. Retos y amenazas

Aplicada a tareas creativas, la IA ha logrado que los ordenadores pasen de ser una herramienta de apoyo a convertirse en el elemento protagonista. En cuestión de segundos o minutos, un algoritmo puede generar contenido original —texto, imágenes, locuciones o vídeos— a partir de unas simples indicaciones, una descripción o unos parámetros determinados. No obstante, cabe insistir en que a medida que la tecnología mejora, también crecen los potenciales aspectos negativos.

Las posibilidades que ofrece la IA son sorprendentes y prometedoras, pero ello mismo las convierte en un arma de doble filo. Es tan sofisticado el logro tecnológico y son tan creíbles los resultados que, sin un código de conducta, la sociedad puede quedar expuesta a formas nuevas de engaño, entre otras amenazas. Ante las conocidas dificultades que

ya experimenta la población general para desenmascarar un bulo o un engaño, incluso con evidencias fáciles de verificar, el panorama que dibuja la creciente sofisticación de los contenidos generados por IA no invita precisamente al optimismo. Es necesario, pues, aumentar la conciencia pública sobre el mal uso potencial de esta tecnología.

### 6.1. La inteligencia artificial toma la palabra

Si bien, como se ha comentado, la simulación de la redacción humana es una gran ayuda para acelerar la producción de textos y mejorar la calidad de la escritura, también tiene sus riesgos, puesto que puede resultar contrapruducente, incluso perjudicial. Si se usa para fines maliciosos, un generador de textos puede escribir de manera convincente reseñas falsas y noticias engañosas. De ahí que el equipo de *OpenAI* se mostrase reacio, en un principio, a poner su modelo *GPT-3* a disposición pública, ya que la interfaz que da acceso al servicio no logra discernir qué tipo de contenido es dañino y cuál es aceptable.

En el ámbito periodístico, el uso potencial de la IA también plantea dudas, sobre todo en un momento en el que la desinformación, en Internet, se ha convertido en una constante. Avances como *GPT-3* implican que el contenido de las noticias falseadas se puede generar sintéticamente de manera que imiten el estilo y el espíritu de las noticias creadas por humanos. Esta posibilidad puede envilecer aún más las campañas de desinformación en línea, sobre todo si se atiende al hecho ya probado de que las personas no suelen ser capaces de distinguir entre el texto generado por una IA y el texto escrito por una persona (**Kreps**; **McCain**; **Brundage**, 2022).

En el terreno académico, la posibilidad de que una IA redacte un artículo completo plantea nuevas dudas éticas y legales sobre la comunicación científica y desata discusione filosóficas sobre la autoría no humana. Es posible que la publicación científica deba afrontar un futuro invadido de manuscritos impulsados por IA y que, al mismo tiempo, se cuestione o se deprecie el valor de los trabajos realizados por el personal investigador cuando haya que atribuir a una máquina una parte no menor del mérito del trabajo (**Osmanovic-Thunström**, 2022). Mientras tanto, cada vez más empresas tecnológicas se están lucrando con el negocio de la escritura asistida sin considerar siquiera la posibilidad de compensar a los incontables autores cuyos textos han utilizado sin permiso para entrenar a sus algoritmos.

La redacción web también plantea interrogantes. ¿Es aceptable en Internet el contenido escrito por una IA? *Google* lo considera fraudulento (**Southern**, 2022) y así lo señala en sus directrices de calidad, que se oponen al texto generado por procedimientos automáticos. Sin embargo, paradójicamente, no cuenta con medios para identificar el contenido redactado por una IA y, por lo tanto, tampoco puede penalizarlo.

### 6.2. El futuro del simulacro audiovisual

En el terreno audiovisual, los vídeos hipertrucados crean la ilusión de que sus protagonistas hacen o dicen cosas que jamás han hecho o han dicho, o participan en situaciones que jamás se produjeron. El peligro de esta posibilidad, cada vez más accesible, reside en que el realismo

de los vídeos ultrafalsos los hace muy creíbles y, por lo tanto, difíciles de verificar. Aunque sean falsos, tienen aspecto de evidencia y, por lo tanto, poseen una capacidad extraordinaria para modelar la opinión pública. En consecuencia, suponen una amenaza para la imagen de cualquier persona, desde un ciudadano anónimo hasta un político o una celebridad, sobre todo porque la democratización de la tecnología *deepfake* está abriendo a más gente la posibilidad de crear falsificaciones convincentes.

Existe también el temor de que esta tecnología pueda llegar a usarse para aportar falsos testimonios en el ámbito jurídico. Este temor lo alimenta la sensación, quizá injustificada, de que a los jueces de nuestro ordenamiento jurídico les falta suficiente preparación en el ámbito tecnológico. Y aunque es cierto que la tecnología debe ser neutra y que un juicio debe basarse en hechos probados, será igualmente difícil asegurar si un vídeo es una prueba audiovisual real o una mera recreación artificial. El conocimiento jurídico se enfrenta, pues, a un nuevo reto: deberá integrar métodos para sopesar y valorar determinadas pruebas a fin de determinar si pueden acreditarse como auténticas.

Dado que las personas, según se ha comprobado, tienden a sobreestimar su capacidad de detectar engaños (**Lyons** *et al.*, 2021), desde el ámbito de la ingeniería se han de destinar más esfuerzos a crear herramientas proactivas de detección. Interesa encontrar modos de hacerle saber al público que se encuentra, según el caso, ante un contenido sintético y potencialmente malicioso. Entre las diversas soluciones que se han formulado está la posibilidad de incrustar una huella digital persistente en los contenidos generados con IA para señalar que las imágenes son el producto de un proceso generativo y que, por lo tanto, no son reales (**Yu** *et al.*, 2021).

En el terreno del entretenimiento y la publicidad, la IA está arrasando de tal forma que se hace necesario un serio debate ético sobre el uso de imágenes de personas fallecidas. El ya mencionado caso de Lola Flores en la campaña de Cruzcampo cosechó reacciones positivas, incluso por parte de la propia familia, porque la agencia trabajó en estrecha colaboración con las hijas de la bailaora. Pero cabe preguntarse si es ético utilizar la imagen de una persona muerta con fines publicitarios. ¿Cómo tener la certeza de que esa persona estaría conforme? ¿Qué ocurriría si mañana se anunciase una bebida alcohólica con un «resucitado» Pau Donés?

### 6.3. La voz humana, seña de identidad

La voz humana sintetizada artificialmente se presta a consideraciones especiales que van más allá de lo tecnológico. Inquietudes como estas, por ejemplo, atrajeron la atención pública durante la producción de *Roadrunner* (2021), un documental sobre el chef Anthony Bourdain, fallecido en 2018. Para lograr que Bourdain pronunciase tres líneas de las que solo había constancia escrita, y para que así aquellas palabras cobrasen vida, los realizadores usaron una versión sintética de su voz. Lo que siguió fue un debate encendido: en contra de las afirmaciones del director, la esposa de Bourdain insistió en que jamás había autorizado aquel uso de su voz (**Dale**, 2022).

Este caso pone de manifiesto la medida en que la voz es una seña de identidad única de cada persona. Por esta razón, el uso de voces clonadas

se debe hacer igualmente en un marco de responsabilidad. Como respuesta a esta necesidad, las empresas de síntesis de voz han introducido pautas éticas sobre el uso de la tecnología y están incorporando medios para garantizar que la persona propietaria de la voz sintetizada realmente ha dado su consentimiento. El propósito de estas medidas es proteger los intereses de quienes prestan su voz, para ayudarles a mantener el control sobre su uso y cobrar unas regalías justas.

### 6.4. Las imágenes sintéticas: un claroscuro

Por otra parte, la capacidad de la IA para interpretar nuestras palabras y crear sobre ellas imágenes artísticas, si bien tiene una cara divertida y amistosa, también esconde un reverso tenebroso. Como el modelo de aprendizaje automático se entrena con datos sin filtrar, los sesgos presentes en ellos se reproducen luego en los resultados. A poco de liberarse, *DALL·E mini* fue objeto de crítica porque las imágenes que arroja representan una cosmovisión marcadamente occidental que además contribuyen a perpetuar y amplificar sesgos raciales y de género (**Pascual**, 2022).

A ello se suma otro frente preocupante: algunos periódicos digitales ya echan mano de *DALL·E mini* y *Craiyon* para ilustrar sus titulares con imágenes falsas (figura 5). Es una práctica reprobable y peligrosa que erosiona la confianza en las evidencias gráficas y allana el camino para nuevas formas de desinformación. Además, a medida que aumente la disponibilidad de modelos más sofisticados y fotorrealistas, tales como *DALL·E 2* o *Imagen*, esta mala praxis supondrá una amenaza aún mayor. Los modelos de IA tienen valor periodístico por su capacidad para producir ilustraciones conceptuales, pero no es aceptable emplearlos para generar contenido fotorrealista que podría confundirse con evidencias fotográficas. Conviene, pues, que se refuerce la deontología periodística en torno a este aspecto y que los agregadores de noticias (*Google Discover*, *Yahoo Noticias*, etc.) la hagan cumplir.


Figura 5. *NextShark* ha sido uno de los primeros medios en ilustrar noticias con imágenes sintéticas (y, por tanto, irreales) sin identificarlas siquiera como tales.

Desde el punto de vista creativo, la popularización de las redes neuronales abre la puerta a nuevas posibilidades artísticas y ya son muchos los creadores que se han aliado con esta nueva tecnología, con la que ahora obtienen inspiración y desarrollan una nueva identidad plástica. Al mismo tiempo, la IA supone una amenaza para los creadores profesionales de contenido visual, que pueden ver devaluada su habilidad para ofrecer trabajos personalizados y de calidad (**McClausland; Salgado**, 2022).

De hecho, ya parece evidente que los ilustradores podrían estar entre los primeros artistas en ser desplazados por la IA (**Barandy**, 2022). Conduce a creerlo así la situación provocada por el escritor Charlie Warzel, colaborador de *The Atlantic*, influido por lo que ya hacían otros autores. Para ilustrar un artículo sobre el teórico de la conspiración Alex Jones, en lugar de adquirir material de un banco de imágenes, optó por producir con *Midjourney* una imagen sintética con toques artísticos. Y fue así como, al reemplazar el trabajo que se le solía encargar a un ilustrador, desató un aluvión de críticas, a las que respondió después en el mismo periódico, en un artículo de reflexión donde admitía su error y lamentaba haber contribuido a sentar tal precedente (**Warzel**, 2022).

A esta controversia se suma la polémica levantada por la Feria Estatal de Colorado al otorgar el primer premio de un certamen de arte a una creación generada mediante una IA (**Roose**, 2022), lo que aviva el temor de que ciertos oficios creativos se vuelvan obsoletos y evidencia cómo la IA está enredando el ya complicado mundo artístico. Además, dado que el aprendizaje automático se nutre del contenido accesible en línea, los artistas que han publicado su obra en Internet pueden haber contribuido, sin saberlo, a entrenar a sus competidores algorítmicos con los que cualquiera, sin formación artística y sin apenas esfuerzo, puede ahora crear elaboradas ilustraciones.

En efecto, hay dudas sobre la licitud y la legalidad de alimentar los modelos de IA con contenido protegido por derechos de autor. *DALL·E 2* y otros generadores de arte producen material gráfico a partir de millones de obras que, en su mayoría, son objeto de propiedad intelectual. ¿A quién pertenece entonces el arte creado a partir de las creaciones y el estilo de otras personas? ¿Y cuánto del trabajo original del artista o fotógrafo es lícito encontrar en las imágenes que genera la IA? ¿No es, en cualquier caso, un robo masivo de propiedad intelectual?

Estos y otros interrogantes surgen de la turbidez de un terreno de juego cuyas reglas todavía no están establecidas, ni menos consensuadas. Es un terreno que aún requiere mayor claridad legal, sobre todo desde que estas producciones se han puesto a la venta y generan ganancias (**Dean**, 2022). Tal es la incertidumbre sobre la autoría de las creaciones sintéticas y la inseguridad jurídica que se deriva que algunos bancos de imágenes, tanto comerciales como gratuitos, se han apresurado a actualizar sus directrices para prohibir cualquier material surgido de un proceso generativo (**Growcoot**, 2022).

### 6.5. Hacia una ética de la inteligencia artificial

Son numerosos, pues, los desafíos que plantea este nuevo gran logro de la tecnología, sobre el cual, por cierto, no hay todavía una comprensión generalizada. Aunque sus aplicaciones se están democratizando a pasos agigantados, a la sociedad en general aún le resulta difícil entender

qué es la IA y de qué es realmente capaz. Lo ilustra el reciente caso del ingeniero Blake Lemoine, despedido de *Google*, tras trabajar en el modelo de lenguaje LaMDA, por sugerir que esta tecnología es consciente y sensible (**Bender**, 2022), pese a que es un hecho aceptado que el potencial de la IA actual, por enorme que sea, no les da a las máquinas la capacidad de pensar o sentir (**Du Sautoy**, 2020; **Metz**, 2022).

Ciertamente, hay propensión a atribuir a la IA unas habilidades cognitivas que no tiene, igualándolas a la representación que de ella se refleja en películas y relatos de ciencia ficción. Pero esta tendencia no es más que una ilusión de profundidad explicativa, un sesgo cognitivo que ha deformado la percepción pública de la IA hasta proyectar de ella, en no pocas ocasiones, una visión exagerada, antropomórfica y deificada.

Aun así, la IA demuestra ser una buena ayuda para el desarrollo rápido y eficiente de un cierto número de tareas que el ser humano realiza de manera menos eficiente. Se avecinan tiempos de transformación en los que podrán verse unas formas novedosas de producir todo tipo de contenidos. Las posibilidades son prometedoras y despiertan curiosidad e interés, pero también causan temor, recelos y preocupación.

En efecto, por una parte hay incógnitas sobre cómo la IA podría sacudir la industria de contenidos o incluso apoderarse de ella, y sobre qué trato merecen los millones de creadores con cuyas obras se alimentan los sistemas de aprendizaje automático. Por otra parte, la naturaleza humana, como es bien sabido, tiene sus flaquezas y puede propiciar usos reprobables, imprudentes o maliciosos de esta nueva tecnología. En consecuencia, las organizaciones implicadas en el desarrollo de aplicaciones de IA deben actuar con transparencia, deben velar por un uso ético y responsable de esta potente herramienta, y deben mantener una estrecha vigilancia para mitigar posibles efectos negativos.

Este trabajo es resultado del proyecto I+D+i *La educación mediática y la dieta informativa como indicadores de la capacidad de análisis crítico de contenidos informativos en futuros docentes* (MEDIA4Teach, PID2019-107748RB-I00/AEI/10.13039/501100011033), que ha contado con la financiación del Ministerio de Ciencia e Innovación del Gobierno de España.

**Referencias**

**Barandy**, **Kat** (2022). «Will artists be replaced by artificial intelligence?». *Designboom*, 10 de agosto. https://designboom.com/art/sebastian-errazuriz-artificial-intelligence-ai-dall-e-replace-artists-illustrators-08-10-2022

**Bender**, **Emily M.** (2022). «Human-like programs abuse our empathy: even Google engineers aren't immune». *The Guardian*, 14 de junio. https://theguardian.com/commentisfree/2022/jun/14/human-like-programs-abuse-our-empathy-even-google-engineers-arent-immune

**Bhargava**, **Cherry**; **Sharma**, **Pardeep Kumar** (eds.) (2022). *Artificial intelligence: fundamentals and applications*. Boca Ratón, FL: CRC Press.

**Boden**, **Margaret A.** (2018). *Artificial intelligence: a very short introduction*. Oxford: Oxford University Press.

**Campesato**, **Oswald** (2020). *Artificial intelligence, machine learning and deep learning*. Dulles, VA: Mercury Learning and Information.

**Dale, Robert** (2022). «The voice synthesis business: 2022 update». *Natural Language Engineering*, v. 28, n. 3, p. 401–408. https://doi.org/10.1017/S1351324922000146

**De-la-Cuadra-de-Colmenares, Elena; López-de-Solís, Iris; Nuño-Moral, María-Victoria** (2014). «Uso de imágenes de archivo en publicidad audiovisual: estudio de casos». *El profesional de la información*, enero-febrero, v. 23, n. 1, pp. 26–35. https://doi.org/10.3145/epi.2014.ene.03

**Dean, Ian** (2022). «You can now sell your DALL·E 2 art, but it feels murky». *Creative Bloq*, 11 de agosto. https://creativebloq.com/news/sell-your-dalle-2

**Du Sautoy, Marcus** (2020). *Programados para crear: cómo está aprendiendo a escribir, pintar y pensar la inteligencia artificial*. Barcelona: Acantilado.

**Dzieza, Josh** (2022). «The great fiction of AI: the strange world of high-speed semi-automated genre fiction». *The Verge*, 20 de julio. https://theverge.com/c/23194235

**Franganillo, Jorge** (2022). *Formatos digitales: propiedades técnicas y contextos de uso*. Barcelona: UOC.

**Growcoot, Matt** (2022). «Getty Images ban AI-generated pictures, Shutterstock following suit». *PetaPixel*, 21 de septiembre. https://petapixel.com/2022/09/21/getty-images-ban-ai-generated-pictures-shutterstock-following-suit

**Kreps, Sarah; McCain, R. Miles; Brundage, Miles** (2022). «All the news that's fit to fabricate: AI-generated text as a tool of media misinformation». *Journal of Experimental Political Science*, v. 9, n. 1, p. 104–117. https://doi.org/10.1017/XPS.2020.37

**Lyons, Benjamin A.; Montgomery, Jacob M.; Guess, Andrew M.; Nyhan, Brendan; Reifler, Jason** (2021). «Overconfidence in news judgments is associated with false news susceptibility». *PNAS*, v. 118, n. 23. https://doi.org/10.1073/pnas.2019527118

**McCausland, Elisa; Salgado, Diego** (2022). «¿Sueñan los androides con ovejas eléctricas? Este es el arte que ya está creando la IA». *El Grito*, 10 de agosto. https://elconfidencial.com/el-grito/2022-08-10/arte-creando-inteligenci-artificial_3473503

**Metz, Cade** (2022). «AI is not sentient: why do people say it is?». *The New York Times*, 5 de agosto. https://nytimes.com/2022/08/05/technology/ai-sentient-google.html

**Nightingale, Sophie J.; Farid, Hany** (2022). «AI-synthesized faces are indistinguishable from real faces and more trustworthy». *PNAS*, v. 119, n. 8. https://doi.org/10.1073/pnas.2120481119

**Nogales, Elena G.** (2022). «En BNElab hemos estado probando el modelo GPT-3 […]». *Twitter*, 27 de junio. https://twitter.com/esnogales/status/1541455846915416066

**Osmanovic-Thunström, Almira** (2022). «We asked GPT-3 to write an academic paper about itself, then we tried to get it published». *Scientific American*, 30 de junio. https://scientificamerican.com/article/we-asked-gpt-3-to-write-an-academic-paper-about-itself-mdash-then-we-tried-to-get-it-published

**Pascual, Manuel G.** (2022) «*Dall·E mini*, el popular generador automático de imágenes que hace dibujos sexistas y racistas». *El País*, 30 de junio. https://elpais.com/tecnologia/2022-06-30/dall-e-el-popular-generador-automatico-de-imagenes-que-hace-dibujos-sexistas-y-racistas.html

**Roose, Kevin** (2022). «An AI-generated picture won an art prize: artists aren't happy». *The New York Times*, 2 de septiembre. https://nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html

**Southern, Matt G.** (2022). «Google says AI generated content is against guidelines». *Search Engine Journal*. https://searchenginejournal.com/google-says-ai-generated-content-is-against-guidelines

**Véliz, Carissa** (2022). «Inteligencia artificial, ¿para qué?». *El País*, 26 de junio. https://elpais.com/eps/2022-06-26/inteligencia-artificial-para-que.html

**Warzel, Charlie** (2022). «I went viral in the bad way: a few lessons from my mistake». *The Atlantic*, 17 de agosto. https://newsletters.theatlantic.com/galaxy-brain/62fc502abcbd490021afea1e/twitter-viral-outrage-ai-art

**Welsh, Shelley** (2022). «The best 10 AI writers & content generators compared». *Search Engine Journal*, 10 de mayo. https://searchenginejournal.com/ai-writers-content-generators

**Yu, Ning; Skripniuk, Vladislav; Abdelnabi, Sahar; Fritz, Mario** (2021). «Artificial fingerprinting for generative models: rooting deepfake attribution in training data». *Proceedings of the IEEE/CVF International Conference on Computer Vision*, p. 14.448–14.457. https://doi.org/10.1109/ICCV48922.2021.01418