

El uso de modelos generativos en el periodismo escrito y audiovisual: un nuevo paradigma en la creación de contenidos

Jorge Franganillo

Universidad de Barcelona

1. INTRODUCCIÓN

La inteligencia artificial (IA) es una rama de la ciencia y la tecnología que busca crear sistemas capaces de resolver tareas que normalmente exigen inteligencia humana. En los últimos años, esta disciplina ha experimentado un gran avance gracias a tres factores: la evolución de los algoritmos, el aumento de la capacidad de cómputo y el acceso a grandes volúmenes de datos. Estos factores han permitido que la IA se utilice en ámbitos tan variados como la medicina, las finanzas, la meteorología o el transporte, entre otros (Boden, 2018; Bhargava y Sharma, 2022).

Una de las aplicaciones de la IA que más interés y debate han generado en la sociedad es la destinada a la creación automatizada de contenido. Para ello se suelen emplear redes neuronales que se entrenan mediante procesos de aprendizaje automático (Campeato, 2020). Estos procesos incluyen el aprendizaje profundo, cuya arquitectura simula la forma en que las personas aprenden habilidades como la predicción de palabras o el reconocimiento de formas.

La IA generativa ha avanzado mucho en capacidad tecnológica y ha alcanzado un nivel de sofisticación sorprendente. Pero conviene tener presente que la seguridad que puede ofrecer una herramienta depende del uso que se le dé. Y como la creatividad humana no tiene límites, la misma tecnología que puede solucionar viejos problemas puede crear problemas nuevos si no recibe un uso apropiado. Este temor justifica

que en los últimos años hayan aparecido iniciativas con impacto social, tales como el proyecto OpenAI, el observatorio OdiseIA o el programa AI for Social Good, de Google, que promueven un uso responsable y ético de la IA para el bien común.

La creación automatizada de contenido ofrece diversas aplicaciones potenciales en los medios de comunicación. Sin apenas intervención humana, la IA generativa puede producir noticias, reportajes, guiones, traducciones y resúmenes, y puede también recrear la imagen y la voz de figuras públicas. Pero, paralelamente, el contenido automático también plantea importantes desafíos éticos, legales y sociales relacionados con la veracidad, la autoría, la responsabilidad o el impacto en la opinión pública. Los recientes avances en IA dibujan, en consecuencia, un panorama complejo que presenta indudables oportunidades, pero también ciertas limitaciones y amenazas.

Ante la magnitud de los progresos orientados a la producción automatizada de contenidos, resulta conveniente examinar estos nuevos desarrollos y los efectos que pueden ocasionar en el público receptor. Siguiendo la línea de investigación iniciada en un artículo anterior (Franganillo, 2022) sobre la industria de contenidos, este trabajo se centra en el ámbito específico del periodismo escrito y audiovisual, sobre el que presentaremos una actualización basada en los últimos avances y ejemplos, y analiza el estado de la cuestión de los modelos generativos aplicados a los medios de comunicación, así como las ventajas y los riesgos que implican para el periodismo escrito y audiovisual, el entretenimiento informativo y la publicidad.

2. LA CREACIÓN DE CONTENIDO, BAJO NUEVAS REGLAS

El aprendizaje profundo es una técnica clave en la IA generativa y se ha convertido en una herramienta valiosa para la producción de contenido en los medios de comunicación. Ha demostrado ser útil en ámbitos como el periodismo escrito, la creación gráfica, la producción audiovisual y el entretenimiento, entre otros. Permite generar textos, voces e imágenes (fijas y en movimiento) de forma cada vez más automatizada,

lo que simplifica tareas, acorta tiempos de producción y mejora así la eficiencia.

Sin embargo, estos avances también plantean importantes cuestiones éticas. Entre ellas se encuentran la transparencia y la responsabilidad en el uso de esta tecnología, el riesgo de contenidos engañosos o sesgados, la protección de la privacidad y de los derechos de autor, y la necesidad de evitar que el uso de la IA provoque discriminación o acentúe desigualdades. Es fundamental abordar estas cuestiones éticas para fomentar en torno a esta tecnología un uso responsable que respete los valores fundamentales de la sociedad.

2.1. Generación de lenguaje natural

El aprendizaje automático tiene muchas aplicaciones prometedoras; entre ellas, la de producir textos que simulan la redacción humana. Con este fin se han creado varios modelos de lenguaje. El más potente hasta ahora es el denominado Generative Pre-trained Transformer, más conocido por sus siglas: GPT. Este modelo, creado por el laboratorio OpenAI, produce textos que imitan de forma convincente el estilo, el tono y el razonamiento de la escritura humana. Y tal es la calidad de los contenidos que genera que resulta difícil distinguirlos de los escritos por una persona (Walsh, 2022).

Pero GPT no solo produce textos, sino que también los adapta, los traduce y los resume basándose en el estudio del contexto, lo que confiere a este modelo un enorme potencial y numerosas aplicaciones prácticas. Ya antes de *ChatGPT* muchas empresas habían comercializado herramientas apoyadas en GPT, tales como *Copy.ai*, *Jasper* o *Wordtune*, para la elaboración automática de textos. Hoy existen aplicaciones específicas en campos como el 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 periodístico, los modelos de lenguaje se usan desde hace tiempo para convertir datos en noticias narradas, de forma casi autónoma. La producción automatizada de noticias, que empezó como un

pequeño experimento, se está extendiendo, aunque con lentitud, a escala global. En cualquier caso, la IA permite hoy producir noticias sobre economía, deportes y meteorología, entre otros temas, y puede ayudar a los redactores, por ejemplo, sugiriendo temas de interés basados en la detección de tendencias.

Algunos medios ya han exhibido el potencial de esta tecnología. La demostración más notoria fue la del diario británico *The Guardian*, que publicó una nota de opinión a partir de una premisa sobre la que GPT-3 generó ocho propuestas (OpenAI, 2020). Los redactores editaron las propuestas como harían con cualquier otro artículo, recortando y reorganizando frases, pero dedicando menos tiempo. Aquel experimento, que otros medios han imitado, sugiere cómo podría ser el periodismo escrito en el futuro.

Ese futuro, de hecho, ya está aquí, pues un buen número de periódicos (*Express*, *Mirror*, *Wall Street Journal*, etc.) y agencias de noticias (Agence France-Presse, Associated Press, Norsk Telegrambyrå, Reuters, etc.) usan la generación de lenguaje natural para automatizar la redacción de artículos (Fanta, 2017; Sánchez, 2021). Producen, sobre todo, noticias basadas en datos cuantitativos como, por ejemplo, los vaivenes de la bolsa, el IPC, la meteorología, alertas de terremotos o resultados deportivos. Y redactan también listículos (Sweney, 2023), que son artículos cuyo contenido se organiza en forma de lista y en párrafos cortos para facilitar la lectura.

Esta automatización presenta una ventaja evidente: en teoría, puede liberar tiempo para el periodismo de investigación y para la elaboración de artículos de fondo, o sea, para tareas de más alto valor. Pero, al menos hoy, los artículos redactados por un modelo suelen ser textos asépticos y algo superficiales que carecen de la puesta en contexto, la narrativa y el comentario de una persona experta en el ámbito. Por tanto, la escritura automática es una práctica que, si se generaliza, podría degradar los contenidos y empobrecer, en consecuencia, la dieta informativa del público.

Al mismo tiempo, hay dudas sobre la licitud y la legalidad de alimentar los modelos con obras que, en su mayoría, tienen derechos de autor. Para entrenar los sistemas de aprendizaje profundo se utiliza el raspado de datos (*web scraping*), un proceso que consiste en recopilar, de manera automática, enormes cantidades de información que está accesible de forma pública. En la Unión Europea, el *Reglamento General de Protección de Datos* no prohíbe ese método, pero sí restringe lo que se puede hacer con los datos extraídos.

El uso de la información recogida puede suponer un problema cuando sustituye la prestación de un servicio. La generación de lenguaje natural está empezando a suplir labores de redacción y edición hasta el momento reservadas a los humanos, y esto resulta preocupante, sobre todo por las posibilidades que abre a la proliferación de contenido engañoso, falso o de mala calidad.

Este temor es un riesgo real: en abril de 2023 la organización News-guard, que evalúa la fiabilidad de las fuentes de información en Internet, encontró medio centenar de sitios de noticias creados casi enteramente por modelos de lenguaje (Sadeghi y Arvanitis, 2023). Se trata, pues, de una nueva ola una nueva ola de granjas de contenido que llenan de anuncios sus artículos *clickbait* para maximizar los ingresos publicitarios. Son sitios de mala calidad que imitan el lenguaje y el aspecto de los medios tradicionales, pero sin ningún editor al frente. Los hay que difunden cientos de artículos al día, algunos con narrativas falsas e información errónea. Y casi todo el contenido presenta un lenguaje superficial y frases repetidas, rasgos típicos del texto generado por IA.

2.2. Síntesis de imágenes a partir de texto

A lo largo de 2022 se popularizó la creación de imágenes a partir de texto gracias a herramientas como *Craiyon*, *DALL·E*, *Midjourney* y *Stable Diffusion*. Estas aplicaciones pueden crear imágenes originales de cualquier tipo, desde objetos y animales hasta paisajes y personas, con solo unas palabras en lenguaje natural. Para ello, emplean redes generativas antagónicas, que son sistemas de IA que se entrenan mediante

el aprendizaje profundo a partir de grandes cantidades de datos. Este logro tecnológico tiene una cara amistosa y hasta divertida, dado que permite producir imágenes sorprendentes con suma facilidad, pero también esconde problemas y riesgos que conviene observar.

Por un lado, al entrenarse con datos sin filtrar, los modelos generativos pueden reproducir sesgos raciales y de género, lo que implica que los algoritmos no son neutrales y que conviene corregir sus desviaciones para garantizar el principio de justicia. Por otro lado, algunas cabeceras usan estas aplicaciones para ilustrar sus titulares con imágenes falsas, pero muy realistas, lo que constituye una práctica peligrosa porque debilita la confianza que se tiene en las evidencias gráficas y puede propiciar nuevas formas de engaño. En efecto, el contenido gráfico falso puede parecer auténtico, como así lo han demostrado algunas imágenes sintéticas divulgadas recientemente, que mostraban escenas falsas del expresidente Donald Trump forcejeando con la policía, del Papa Francisco con un abrigo de plumas y de un abrazo entre los políticos Yolanda Díaz y Pablo Iglesias (imagen 1).

Imagen 1. El falso abrazo de Yolanda Díaz y Pablo Iglesias abrió una portada impresa.



Fuente: *El Mundo*.

Desde otra óptica, la IA generativa puede tener valor periodístico porque permite crear ilustraciones abstractas y conceptuales. Es una tecnología al alcance de cualquiera, sin necesidad de formación en diseño gráfico, lo que también la hace peligrosa. Aunque el contenido generado no vulnere la imagen de nadie, puede perjudicar a profesionales de la fotografía y la ilustración, que pueden ver desplazada y devaluada su habilidad para ofrecer obras personalizadas y de buena calidad (Barandy, 2022; McCausland y Salgado, 2022).

Los usos editoriales de la síntesis gráfica son cada vez más habituales. Desde junio de 2022 hemos visto varios ejemplos que han despertado tanto admiración como recelo. La revista británica *The Economist* usó *Midjourney* para ilustrar la portada de un número dedicado a las promesas y los peligros de la IA. Poco después, la revista *Cosmopolitan* demostró las capacidades de esta tecnología, usando *DALL·E 2*. Pero fue polémico el caso de Charlie Warzel, un colaborador de la revista *The Atlantic*, muy criticado por ilustrar un artículo de opinión con una imagen sintética. Al hacerlo, prescindió del trabajo de un ilustrador y del material de un banco de imágenes, y sentó un precedente que él mismo lamentó después. Estos hitos muestran la rapidez con que las tecnologías digitales pasan de la innovación a la mercantilización, y cómo en el proceso surgen nuevos miedos.

2.3. Vídeos hipertrucados: tecnología *deepfake*

En el campo audiovisual, la IA generativa ha conquistado espacios televisivos a través de la tecnología *deepfake*, que es capaz de generar vídeos manipulados con un aspecto tan realista que hace difícil notar que han sido falseados. Estos contenidos, que no son reales, pero que lo parecen gracias a una manipulación extrema para hacer creer que una persona hace o dice algo que no es real, se denominan también *ultra-falsos* o *hipertrucados*, dos términos que reflejan con acierto la naturaleza de estas imágenes. Estos adjetivos implican que las imágenes se pueden falsear de forma tan radical, tal es la sofisticación de esta tecnología, que apenas se percibe el engaño.

El cine fue pionero en recrear personas de forma realista mediante el CGI (siglas en inglés de «imágenes generadas por ordenador»), una técnica que permitía sustituir digitalmente el rostro de artistas fallecidos por el de un doble. Otra innovación cinematográfica fue el «efecto Forrest Gump», que se originó en la película homónima de 1994. Este truco consiste en mezclar elementos actuales e históricos en las imágenes, y ha sido usado después por marcas como Peugeot, Mercedes Benz o Virgin Trains en sus anuncios. Sin embargo, estas técnicas tienen el problema de ser costosas, ya que requieren personal especializado y mucho tiempo de trabajo.

La tecnología *deepfake* ha revolucionado la recreación de personas en las industrias creativas y ha abierto nuevas posibilidades para el cine y la publicidad. Ha permitido rejuvenecer a personajes de ficción, como los interpretados por Robert De Niro en *El irlandés* (2019), o Harrison Ford en *Indiana Jones y el dial del destino* (2023). Y ha logrado «revivir» a personalidades fallecidas, como Lola Flores, en la campaña «Con mucho acento» (2021) de Cruzcampo, o Salvador Dalí, en campañas del Museo Dalí de Florida (2019) y de la Fundación Reina Sofía (2021).

Permite incluso suplantar en tiempo real a figuras públicas con fines satíricos, como lo ha mostrado *El Intermedio* (La Sexta) en las secciones «Entrevista por la cara» y «Aznarito y Felipón», donde el presentador Gran Wyoming y su colaborador Dani Mateo encarnan a figuras políticas reconocidas. Esta misma aplicación de la IA ha impulsado la creación y difusión de vídeos más o menos divertidos en los que el rostro de estrellas como Steve Buscemi o Nicolas Cage suplantan el de personajes de cualquier película. Y también ha popularizado canales de *TikTok* que triunfan con imitadores que suplantan a celebridades como Tom Cruise, Keanu Reeves o Margot Robbie.

La falsificación de contenido audiovisual no es un fenómeno nuevo. La novedad es lo mucho que la IA ha facilitado el proceso. Antes se necesitaba la intervención humana para realizar el trabajo manual y tedioso de manipular las imágenes, pero ahora se puede entrenar a una

IA para que superponga una cara en un vídeo de forma autónoma. Los resultados son tan sofisticados que la ficción y la realidad se confunden.

Esto ha generado controversia en el ámbito del entretenimiento informativo. En mayo de 2022, la televisión pública francesa estrenó *Hôtel du temps*, un formato que utilizaba la tecnología *deepfake* para recrear a celebridades fallecidas y hacerles hablar con el presentador, Thierry Ardisson, rejuvenecido artificialmente (imagen 2). Entre las figuras invitadas se encuentran Lady Di, François Mitterrand y Dalida, la diva de la canción francesa a quien el presentador le pregunta nada menos que por su suicidio. Ante esta situación, surge la cuestión ética de si es correcto usar una recreación tan verosímil de estas personas diciendo cosas que quizá nunca pronunciaron. ¿Cómo podemos tener la certeza de que estarían conformes?

Imagen 2. Un instante del programa *Hôtel du temps* (2022).



Fuente: France 3.

Además de la falta de consentimiento de las personas recreadas por IA, existe el dilema ético de que estas falsificaciones se usen para generar bulos, fraudes o ataques a la reputación de personas, organizaciones o gobiernos. Estos contenidos son difíciles de detectar y pueden resultar muy destructivos. Antes se podía notar una sensación de extrañeza ante

la mirada vacía de una persona artificial, lo que se conoce como «efecto del valle inquietante». Pero hoy las imágenes son tan convincentes que las caras sintetizadas no solo se confunden con las reales, sino que además generan más confianza (Nightingale y Farid, 2022).

La tecnología que permite crear contenidos artificiales con algoritmos es entonces un gran avance técnico, pero también una amenaza potencial. Con solo unos clics, cualquiera puede crear falsificaciones sintetizadas sin necesidad de conocimientos especializados o de una gran inversión. Esos contenidos pueden usarse para fines perversos, como el fraude financiero, la difamación política, el chantaje con falsas imágenes íntimas y, por qué no, para nuevas formas de fraude y abuso. La amenaza es real: el 96 % de los vídeos *deepfake* publicados en línea en 2019 eran pornográficos y no consentidos, siendo las mujeres el colectivo más afectado (Ajder *et al.*, 2019). Y en 2020 se identificaron más de 85.000 vídeos dañinos contra la reputación de figuras públicas, creados a un ritmo que se duplicaba cada seis meses (Ajder, 2020).

2.4. Clonación de voz

La síntesis del habla es una tecnología sencilla que permite convertir texto en una voz muy similar a la humana, y que se usa desde hace décadas como herramienta de apoyo en situaciones diversas: asistentes de voz, lectores de pantalla, navegación GPS, aprendizaje de idiomas y atención telefónica automatizada, entre otras. Estos ejemplos muestran que no hace falta una tecnología sofisticada para ofrecer soluciones prácticas.

Actualmente, el aprendizaje profundo ha ampliado las posibilidades de uso comercial de la síntesis de voz (Dale, 2022). Una de ellas es la de clonar voces famosas para usarlas, con ética y con licencia, en audiolibros, productos audiovisuales y entornos inmersivos, como voces en *off* o voces de doblaje. En esta fórmula es pionera la empresa Veritone, cuyo servicio Veritone Voice ofrece voces de celebridades, actores, actrices, deportistas y otras personas influyentes para cualquier proyecto sonoro o multimedia. Una ventaja de usar una IA es que acelera procesos y

recorta gastos. La voz sintetizada se genera con un algoritmo en tiempo real, sin que la persona que la presta deba desplazarse a un estudio y dedicar horas a grabar una locución; solo cobra unos *royalties* por el uso comercial de su voz.

Este uso es un ejemplo de cómo la IA puede resolver tareas expertas sin el coste y las posibles trabas que supone contratar a una persona experta. Esta es, de hecho, la principal característica de los llamados «sistemas expertos», que se desenvuelven como humanos, pero sin las limitaciones ni las lógicas exigencias humanas.

Los medios de comunicación pueden usar la clonación de voces con varios fines. Por ejemplo, para narrar noticias y eventos con la voz de un periodista o presentador famoso, lo que hace más atractiva la información. O para doblar películas y series con las voces del elenco en otro idioma. O incluso para dar voz a personajes históricos o fallecidos en documentales y programas de televisión, cuando ya no se puede grabar su propia voz.

Imagen 3. Un fotograma del documental *Anthony Bourdain: un chef por el mundo* (2021).



Fuente: *IMDb*.

Sin embargo, este uso puede ser polémico, como ya ocurrió con la producción del documental *Anthony Bourdain: un chef por el mundo* (2021), sobre el cocinero estadounidense, que se había suicidado poco antes (imagen 3). Para lograr que Bourdain pronunciara unas palabras que había escrito en vida, los realizadores clonaron su voz, pero sin la aprobación de su viuda. Este incidente levantó mucho debate, pues la voz es una seña de identidad, y falsificarla en un documental, que es una expresión de la realidad, incurre en una manipulación.

Esta tecnología tiene además otros peligros, porque se puede usar con fines maliciosos. Por ejemplo, se podría usar la voz clonada de un periodista, un presentador o una celebridad y hacerles decir cosas falsas o dañinas que afectasen a su reputación y credibilidad. O bien se podrían difundir noticias falsas y desinformación, manipulando al público espectador. Estas prácticas pueden violar el derecho a la privacidad y a la imagen de las personas, y pueden generar desconfianza y confusión en el público. Por eso, es importante regular el uso de la clonación de voz y establecer mecanismos de verificación y transparencia que permitan identificar la fuente y la autenticidad de las grabaciones sonoras.

3. RETOS Y AMENAZAS

Con la irrupción de la IA generativa en las tareas creativas, los ordenadores han dejado de ser herramientas de apoyo para pasar a ser protagonistas, capaces de crear contenido original. Con unas indicaciones sencillas, una descripción simple o unos parámetros básicos, un algoritmo puede generar contenido original —texto, imágenes, vídeos o locuciones— en cuestión de segundos o minutos.

La tecnología avanza inexorable y todo apunta a que los modelos mejorarán la capacidad para comprender el lenguaje y para generar contenido variado. A pesar de sus flaquezas, tendemos hacia una adopción generalizada de estos sistemas, lo que propiciará que el contenido automático sea cada vez más habitual. De hecho, se están haciendo esfuerzos para favorecer la integración de la IA generativa en toda clase

de entornos, tales como buscadores, navegadores, paquetes ofimáticos y programas de edición de imagen y vídeo.

Pero esta tecnología también tiene sus riesgos; a medida que avanza, aumentan las posibilidades de engaño y otras amenazas. Los contenidos artificiales son tan verosímiles que pueden confundir a la sociedad si no hay un código de conducta. Si para muchas personas ya era difícil separar la verdad de lo engañoso con evidencias simples, el panorama que ahora se presenta con esta creciente sofisticación es desalentador y preocupante. Por ello, es necesario aumentar la conciencia pública sobre el uso responsable y ético de esta tecnología.

3.1. Implicaciones del contenido artificial

La escritura asistida es un gran recurso para agilizar la producción de textos y mejorar la calidad de la expresión, pero también entraña riesgos y puede acarrear consecuencias negativas. En el ámbito periodístico, la IA generativa provoca incertidumbre, sobre todo en estos tiempos en los que la desinformación se ha apoderado de los medios y las redes sociales.

Con la ayuda de un modelo de lenguaje se pueden crear noticias falsas con fines maliciosos y de forma convincente, imitando el estilo de las informaciones redactadas por humanos, como ya hace una nueva generación de granjas de contenido. Incluso en contextos benignos, los modelos de lenguaje pueden generar contenido tóxico, ya que su interfaz no siempre logra discernir qué tipo de contenido es dañino y cuál es aceptable. Esta circunstancia puede recrudecer las campañas de desinformación en línea, sobre todo si se atiende al hecho de que el texto creado por un modelo de lenguaje es casi indistinguible del creado por una persona (Kreps *et al.*, 2022).

Los gigantes de las búsquedas, por su parte, tampoco pueden hacer mucho para detectar textos automáticos y advertir de ello. Google, por ejemplo, considera fraudulento el contenido escrito por una IA (Southern, 2022), y así lo señala en sus directrices de calidad, que se oponen sin atenuantes al texto generado por procesos automáticos. Pero no

tiene medios para identificarlo y, por lo tanto, no puede señalarlo ni tampoco penalizarlo.

Aunque los modelos de lenguaje pueden producir textos naturales y fluidos, tienen ciertas limitaciones. Por ejemplo, no abarcan toda la información accesible en Internet, sino una selección de datos relevantes, que en su mayoría proceden de fuentes de habla inglesa. Además, no siempre se basan en fuentes fiables o evidencias robustas. El texto que generan es una mera construcción sintética cuyo contenido se debe contrastar con fuentes fiables y expertas. La IA generativa solo es válida, entonces, para situaciones que admiten cierto margen de error. No sirve para cuestiones críticas, como un trabajo científico, un consejo legal o financiero, o una consulta médica. Tampoco razona ni entiende, en un sentido humano, nada de lo que escribe. El texto que produce puede ser incorrecto o absurdo, porque la IA se inventa hechos, personas, datos o fuentes. No dispone de un modelo de verdad y puede confundir al público lector.

Preocupa también la práctica cada vez más frecuente entre los medios digitales de ilustrar sus noticias con imágenes sintéticas. Esta tendencia socava la confianza en las evidencias gráficas y aumenta la confusión. El riesgo está aumentando a medida que los modelos logran imágenes más realistas. Por lo tanto, aunque los modelos de IA pueden servir muy bien para crear ilustraciones conceptuales, no son adecuados para simular evidencias fotográficas. Por ello, se necesita reforzar la ética periodística en este aspecto y que los agregadores de noticias la hagan cumplir con rigor profesional.

La verosimilitud del contenido generado por máquinas puede confundir al público si no se explicita su origen, lo cual ha suscitado ya varias polémicas en concursos de arte y fotografía (DigiDirect, Feria Estatal de Colorado, Sony World Photography Awards), que han premiado creaciones sintéticas sin saber que lo eran. Esto aviva el temor de que ciertos oficios creativos se vuelvan obsoletos y evidencia cómo la IA está enredando el ya complicado mundo de lo artístico, que no parece preparado para las implicaciones de esta nueva forma de creación.

Además, como el aprendizaje automático se nutre del contenido accesible en línea, quienes han publicado su obra en Internet pueden haber alimentado, sin proponérselo, a sus rivales algorítmicos, con los que cualquiera puede ahora crear ilustraciones elaboradas sin formación artística ni esfuerzo.

En el ámbito creativo, las redes neuronales son una fuente de inspiración y expresión para muchos artistas, que exploran nuevas posibilidades con esta tecnología. Pero también son una amenaza para los profesionales del contenido visual, que pueden ver su devaluado su trabajo (McClausland y Salgado, 2022). De hecho, ya parece evidente que los ilustradores podrían estar entre los primeros artistas en ser desplazados por la IA (Barandy, 2022). Ante tal riesgo, cabe preguntarse si es lícito que los modelos se alimenten de contenido gráfico protegido. ¿De quién es el contenido que se crea a partir del estilo y las obras de otras personas? ¿Y qué proporción del trabajo original del artista o fotógrafo es aceptable hallar en las imágenes sintéticas?

Son interrogantes que surgen en un ámbito que aún no tiene unas normas claras ni consensuadas. Es un terreno que necesita más claridad legal, sobre todo desde que estas creaciones se han comercializado y generan ganancias (Dean, 2022); las aplicaciones generativas se han vuelto un negocio muy lucrativo. Es tal la incertidumbre que algunos bancos de imágenes han reescrito sus directrices para impedir la venta de materiales surgidos de un proceso generativo (Growcoat, 2022).

Por otra parte, los algoritmos de generación de imágenes, si se entrenan con datos sin filtrar, pueden reproducir los estereotipos culturales, étnicos o de género que haya en ellos. Estos sesgos causaron críticas cuando se descubrió que algunos modelos de visión artificial y de síntesis de imágenes mostraban una visión discriminatoria al incorporar ideas sexistas y racistas presentes en los datos de entrenamiento (Hao, 2021; Steed y Caliskan, 2021). Con el tiempo, no obstante, estas desviaciones se han ido corrigiendo, pero ilustran que los sistemas no solo deben ser competentes, sino que además deben actuar con ética.

Se deben calibrar y reentrenar para que eviten perpetuar sesgos perjudiciales y mejoren así la calidad de los resultados.

En lo audiovisual, los vídeos hipertrucados crean la ilusión de que sus protagonistas hacen o dicen cosas que nunca hicieron o dijeron, o participan en situaciones que nunca ocurrieron. El peligro de esta posibilidad, cada vez más accesible, radica en que el realismo de los vídeos ultrafalsos los hace muy creíbles y difíciles de verificar. Aunque sean falsos, tienen aspecto de evidencia y, por tanto, tienen un gran poder para manipular la opinión pública. Entonces, suponen una amenaza para la imagen de cualquier persona, desde un ciudadano anónimo hasta un político o una celebridad, sobre todo porque la tecnología *deepfake* está al alcance de más gente que puede crear con ella falsificaciones convincentes.

Dado que las personas, como se ha comprobado, tienden a sobreestimar su capacidad de detectar engaños (Lyons *et al.*, 2021), desde el ámbito de la ingeniería se deben dedicar más esfuerzos a crear herramientas proactivas de detección. Interesa encontrar formas de informar al público de que se encuentra, según el caso, ante un contenido sintético y potencialmente malicioso. Para evitar el engaño de los contenidos artificiales, se ha propuesto insertar una huella digital indeleble que indique que las imágenes son fruto de un algoritmo y no de la realidad (Yu *et al.*, 2021).

En el ámbito del entretenimiento y la publicidad, la recreación de imágenes de personas fallecidas exige un serio debate ético, serio y responsable. El caso de Lola Flores en la mencionada campaña de Cruzcampo tuvo buena acogida, incluso por parte de su familia, ya que la agencia colaboró con sus hijas para recrear a la artista. Pero cabe preguntarse si es ético usar la imagen de un difunto para fines comerciales. ¿Cómo saber si esa persona estaría de acuerdo? ¿Qué pasaría si mañana se recrease al desaparecido compositor y cantante Pau Donés, que murió de cáncer, para promocionar, precisamente, una bebida alcohólica?

Por último, la voz humana sintetizada artificialmente también plantea cuestiones que van más allá de lo tecnológico. Cada nueva

producción sonora o audiovisual que recrea la voz de una personalidad conocida atrae la atención pública y genera inquietud. La voz es una señal de identidad única de cada persona, por lo que el uso de voces clonadas debe hacerse en un marco de responsabilidad. Para responder a este imperativo, las empresas de síntesis de voz han creado pautas éticas y han incorporado medios para asegurar que la persona dueña de la voz sintetizada ha dado su consentimiento. Son medidas que buscan proteger los intereses de quienes prestan su voz, para que puedan controlar su uso y recibir unas regalías justas.

3.2. Hacia una ética de la IA

Son muchos los retos que plantea este nuevo gran logro de la tecnología, sobre el cual, por cierto, no hay todavía una comprensión generalizada. Aunque la IA está llegando rápidamente a todos los públicos, no se puede afirmar que se esté democratizando, ya que esos públicos no solo han de poder usarla, sino que también deben poder entenderla. La democratización implica tanto el acceso como la educación, la ética y la transparencia. Por lo tanto, aunque es innegable que la IA se está democratizando en algunos aspectos, también se debe admitir que hay intereses privados que limitan o condicionan ese proceso.

A la sociedad, en general, todavía le resulta difícil entender qué es la IA y de qué es realmente capaz, como así lo ilustra el reciente caso del ingeniero Blake Lemoine, despedido de Google por sugerir que el modelo de lenguaje LaMDA es consciente y sensible (Bender, 2022), aunque bien se sabe que la IA no puede pensar ni sentir (Du-Sautoy, 2020; Metz, 2022). Hay cierta inclinación a atribuir a la IA unas capacidades cognitivas que en realidad no tiene, igualándolas a la representación que de ella se ve en películas y relatos de ciencia ficción. Pero esta tendencia no es más que una ilusión de profundidad explicativa, es decir, un sesgo cognitivo que distorsiona la percepción pública de la IA y a menudo proyecta de ella una visión exagerada, antropomórfica y deificada.

Aun así, la IA ofrece una valiosa ayuda para agilizar una buena cantidad de tareas en las que el ser humano es menos eficiente. Se avecinan tiempos de transformación y nuevas formas de producir contenidos. Las posibilidades son prometedoras y atractivas, pero también generan miedo, recelo y preocupación.

En efecto, hay incógnitas sobre cómo la IA podría sacudir o incluso dominar la industria de contenidos, y sobre qué derechos tienen los creadores cuyas obras nutren el aprendizaje automático. Además, la naturaleza humana puede darle a esta nueva tecnología un uso reprobable, imprudente o malicioso. Por ello, las organizaciones que desarrollan aplicaciones de IA deben ser transparentes, éticas y responsables con esta potente herramienta, y deben mantener una estrecha vigilancia para prevenir, o si fuera el caso compensar, posibles efectos negativos.

El Parlamento Europeo ya ha establecido las bases para regular el uso de los datos que alimentan los modelos generativos y prevenir los usos maliciosos de la IA que puedan vulnerar los derechos de protección de datos o la privacidad de los ciudadanos (Ayuso y Pascual, 2023). Las empresas tecnológicas deberán ser transparentes e indicar claramente cuándo los resultados generados en forma de texto, imágenes o audios son creaciones generativas, para evitar confundir al ciudadano sobre su autoría y su autenticidad. También tendrán que señalar de forma explícita las fuentes de las que se ha nutrido y con las que se ha entrenado al sistema, para proteger la propiedad intelectual. Es crucial legislar este aspecto e incluso prever formas de compensación económica para garantizar la viabilidad de la industria de medios y el ecosistema cultural del que se aprovechan hoy gratuitamente los modelos generativos.

Asimismo, los gobiernos de todo el mundo deberán colaborar con la comunidad científica para crear leyes que protejan los derechos individuales y pongan límites penales al uso tóxico de esta tecnología. Además, deberán lanzar campañas de concienciación para que el público comprenda el alcance de la inteligencia artificial generativa y aprenda a defenderse de las nuevas falsificaciones.

4. REFERENCIAS

- AJDER, H. *Deepfake threat intelligence: a statistics snapshot from June 2020*. Sensity, 2020.
- AJDER, H.; PATRINI, G.; CAVALLI, F.; CULLEN, L. *The state of deepfakes: landscape, threats and impact*. DeepTrace, 2019. <https://bit.ly/3NWoWX7> [Consulta: 30/05/2023.]
- AYUSO, S.; ASCUAL, M.G. «Europa quiere poner más obligaciones a la inteligencia artificial generativa como la de ChatGPT». *El País*, 11 de mayo de 2023. <https://bit.ly/3ryRjDi> [Consulta: 01/06/2023.]
- BARANDY, K. «Will artists be replaced by artificial intelligence?». *Designboom*, 10 de agosto de 2022. <https://bit.ly/3Q3Wkho> [Consulta: 01/06/2023.]
- BENDER, E.M. «Human-like programs abuse our empathy: even Google engineers aren't immune». *The Guardian*, 14 de junio de 2022. <https://bit.ly/3K6c93c> [Consulta: 01/06/2023.]
- BHARGAVA, C.; Sharma, P.K. (eds.). *Artificial intelligence: fundamentals and applications*. Boca Ratón, FL: CRC Press, 2022.
- BODEN, M.A. *Artificial intelligence: a very short introduction*. Oxford: Oxford University Press, 2018.
- CAMPESATO, O. *Artificial intelligence, machine learning and deep learning*. Dulles, VA: Mercury Learning and Information, 2020.
- DEAN, I. «You can now sell your DALL·E 2 art, but it feels murky». *Creative bloq*, 11 de agosto de 2022. <https://bit.ly/44uSeU1> [Consulta: 01/06/2023.]
- DALE, R. «The voice synthesis business: 2022 update». *Natural language engineering*, 28(3), 2022, pp. 401–408.
- DU SAUTOY, M. Programados para crear: cómo está aprendiendo a escribir, pintar y pensar la inteligencia artificial. Acantilado, 2020.
- FANTA, A. *Putting Europe's robots on the map: automated journalism in news agencies*. Oxford: University of Oxford; Reuters Institute for the Study of Journalism, 2017. <https://bit.ly/3K5pk4r> [Consulta: 30/05/2023.]
- FRANGANILLO, J. «Contenido generado por inteligencia artificial: oportunidades y amenazas». *Anuario ThinkEPI*, 16, 2022, e16a24.

- GROWCOOT, M. «Getty Images ban AI-generated pictures, Shutterstock following suit». *PetaPixel*, 21 de septiembre de 2022. <https://bit.ly/3Dj8MlJ> [Consulta: 01/06/2023.]
- KREPS, S.; MCCAIN, R.M.; BRUNDAGE, M. «All the news that's fit to fabricate: AI-generated text as a tool of media misinformation». *Journal of Experimental Political Science*, 9(1), 2022, pp. 104–117.
- LYONS, B.A.; MONTGOMERY, J.M.; GUESS, A.M.; NYHAN, B.; REIFLER, J. «Overconfidence in news judgments is associated with false news susceptibility». *PNAS*, 118(23), e2019527118, 2021.
- MCCAUSLAND, E.; SALGADO, D. «¿Sueñan los androides con ovejas eléctricas? Este es el arte que ya está creando la IA». *El Grito*, 10 de agosto de 2022. <https://bit.ly/3DhSJVk> [Consulta: 01/06/2023.]
- METZ, C. «AI is not sentient: why do people say it is?». *The New York Times*, 5 de agosto de 2022. <https://nyti.ms/3K3AYNt> [Consulta: 01/06/2023.]
- NIGHTINGALE, S.J.; FARID, H. «AI-synthesized faces are indistinguishable from real faces and more trustworthy». *PNAS*, 119(8), e2120481119, 2022.
- OPENAI. «A robot wrote this entire article: are you scared yet, human?». *The Guardian*, 8 de septiembre de 2020. <https://bit.ly/3K2BFGJ> [Consulta: 31/05/2023.]
- OSMANOVIC-THUNSTRÖM, A. «We asked GPT-3 to write an academic paper about itself, then we tried to get it published». *Scientific American*, 30 de junio de 2022. <https://bit.ly/44SF3vB> [Consulta: 01/06/2023.]
- PATRINI, G. The state of deepfakes, 2020: updates on statistics and trends. Sensity, 2021.
- SADEGHI, M.; ARVANITIS, L. «Rise of the newsbots: AI-generated news websites proliferating online». *Newsguard*, 1 de mayo de 2023. <https://bit.ly/3Dm0oBY> [Consulta: 01/06/2023.]
- SÁNCHEZ, S. «The Wall Street Journal uses Narrativa's AI for its news automation». *Narrativa*, 29 de enero de 2021. <https://bit.ly/43wfxv5> [Consulta: 30/05/2023.]
- SOUTHERN, M.G. «Google says AI generated content is against guidelines». *Search Engine Journal*, 6 de abril de 2022. <https://bit.ly/3O61ZkC> [Consulta: 01/06/2023.]

- SWENEY, M. «*Mirror* and *Express* owner publishes first articles written using AI». *The Guardian*, 7 de marzo de 2023. <https://bit.ly/3K6csuS> [Consulta: 01/06/2023.]
- WALSH, S. «The best 10 AI writers & content generators compared». *Search Engine Journal*, 10 de mayo de 2022. <https://bit.ly/3Oj65Gb> [Consulta: 01/06/2023.]
- YU, N.; SKRIPNIUK, V.; ABDELNABI, S.; FRITZ, M. «Artificial fingerprinting for generative models: rooting deepfake attribution in training data». *Proceedings of the IEEE/CVF International conference on computer vision*, 2021, pp. 14.448–14.457.