

La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos

Generative artificial intelligence and its impact on media content creation

Jorge Franganillo  | franganillo@ub.edu
Universidad de Barcelona, España

10.17502/mrcs.v11i2.710

Recibido: 01-06-2023

Aceptado: 11-09-2023



Resumen

La inteligencia artificial (IA) generativa es un campo en rápido avance que ya permite la producción automatizada de contenido textual, gráfico, sonoro y audiovisual de alta calidad. Esta tecnología tiene implicaciones relevantes para el periodismo, la publicidad y el entretenimiento, así como también plantea desafíos éticos, legales y sociales. Este trabajo examina las posibilidades, las limitaciones y los riesgos de la IA generativa para la producción de contenidos en los medios de comunicación. Se analizan los grandes modelos de lenguaje para la generación automatizada de texto, las redes antagónicas generativas para la síntesis de imágenes y vídeos cortos, y la tecnología *deepfake* para la falsificación de vídeos y la clonación de voces humanas. Se discuten las implicaciones de estas tecnologías para la propiedad intelectual, la veracidad informativa, la identidad personal y la creatividad humana. Como conclusión puede afirmarse que la IA generativa es una herramienta potente e innovadora para la creación de contenidos mediáticos, pero que requiere un uso ético y cuidadoso por parte de los productores y de los consumidores de contenido.

Palabras clave: algoritmos creativos, contenidos mediáticos, ética de la tecnología, inteligencia artificial, modelos generativos.

Abstract

Generative artificial intelligence (AI) is a rapidly advancing field that enables the automated production of high-quality textual, graphic, sound and audiovisual content. This technology has significant implications for journalism, advertising and entertainment, as well as ethical, legal and social issues. This paper examines the possibilities, limitations and risks of generative AI for content production in the media. It analyzes large language models oriented to automated writing, generative adversarial networks oriented to image and short video synthesis, and deepfake technology for video manipulation and voice cloning. The implications of these technologies for intellectual property, information veracity, personal identity and human creativity are discussed. It is concluded that generative AI is a powerful and innovative tool for media content creation, but that it also requires careful and ethical use by both content producers and consumers.

Keywords: artificial intelligence, creative algorithms, generative models, media content, technology ethics.

Sumario

1. Introducción | 2. Metodología | 3. Un nuevo paradigma en la creación de contenidos | 3.1. Redacción automatizada de noticias | 3.2. Síntesis de imágenes a partir de texto | 3.3. Vídeos hipertrucados: tecnología *deepfake* | 3.4. Síntesis de vídeo a partir de texto | 3.5. Clonación de voz | 4. Discusión y conclusiones | 4.1. Implicaciones del contenido artificial | 4.1.1. Suplantación de la escritura humana | 4.1.2. Suplantación de la creación gráfica | 4.1.3. Suplantación de la realidad: vídeos sintéticos y ultrafalsos | 4.1.4. Suplantación de la voz humana | 4.2. Hacia una ética de la IA | 4.3. Futuras líneas de investigación | Referencias.

Cómo citar este artículo

Franganillo, J. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *methaodos.revista de ciencias sociales*, 11(2), m231102a10. <http://dx.doi.org/10.17502/mrcs.v11i2.710>

1. Introducción

La inteligencia artificial (IA) es una disciplina científica y tecnológica que busca crear sistemas capaces de resolver tareas que normalmente requieren de inteligencia humana. En los últimos años, esta disciplina ha experimentado un gran avance gracias a tres factores: el desarrollo de algoritmos más sofisticados, el aumento de la capacidad de cómputo y el acceso a enormes cantidades de datos. Estos factores han permitido que la IA se aplique a ámbitos tan diversos como la medicina, las finanzas, la meteorología o el transporte, entre otros (Boden, 2018; Bhargava y Sharma, 2022).

En el desarrollo de la IA se pueden distinguir tres oleadas, que se diferencian según sus capacidades. La primera oleada —la IA simbólica— consiste en programar las reglas lógicas para que un sistema pueda actuar de forma inteligente, como jugar al ajedrez o diagnosticar una enfermedad. La segunda oleada —la IA basada en datos— utiliza técnicas de aprendizaje automático para extraer patrones y conocimiento a partir de grandes cantidades de datos, y es la responsable del auge actual de la IA. La tercera oleada —la IA basada en contextos— es la base de la superinteligencia artificial, un hipotético sistema capaz de comprender el mundo real y de ofrecer soluciones innovadoras a problemas nuevos o desconocidos. Estas oleadas también se diferencian por el grado de complejidad y autonomía de la IA, que así se puede clasificar en IA específica o débil, e IA general o fuerte. Hasta ahora, solamente se han desarrollado formas de IA específica, que solo pueden resolver tareas concretas y limitadas. La IA general, por su parte, es un objetivo muy ambicioso y complejo, que suscita opiniones divergentes sobre si es realmente posible y sobre todo deseable lograrla (Boucher, 2020).

Una de las aplicaciones de la IA que más ha permeado socialmente y que más impacto y controversia ha causado es la destinada a la creación automatizada de contenido. Para ello se suelen usar redes neuronales que se entrenan mediante procesos de aprendizaje automático (Campesato, 2020). Estos procesos incluyen el aprendizaje profundo, un tipo de aprendizaje automático cuya arquitectura imita la forma en que las personas aprenden habilidades como, por ejemplo, la predicción de palabras o el reconocimiento de formas. Los modelos generativos procesan un gran corpus de datos complejos y no estructurados, tales como textos, audios o imágenes, para luego generar contenido nuevo con el mismo estilo que los datos originales.

La IA generativa ha acelerado su crecimiento en capacidad tecnológica y está alcanzando un grado de sofisticación antes difícil de imaginar. Pero es necesario tener en cuenta que la seguridad que puede ofrecer esta herramienta, como ya se sabe, depende del uso que se le dé. Sabiendo que la creatividad humana no tiene límites, ni parece que los vaya a tener, la misma tecnología capaz de resolver viejos problemas puede provocar problemas nuevos si no recibe un uso apropiado. De ahí que en los últimos años hayan surgido iniciativas con impacto social, tales como el proyecto OpenAI, el observatorio OdiselAI o el programa AI for Social Good, de Google, que promueven el uso ético y responsable, y siempre para el bien común, de la IA.

El uso de la IA generativa para la creación de contenidos plantea implicaciones en diversos ámbitos, como el laboral, el económico y el educativo. En cuanto al mercado laboral, por ejemplo, se estima que el impacto global de la IA no será significativo hasta más allá del año 2030. Así lo indicaba un informe de PwC (2018), que entonces señalaba que solo el 3 % de los empleos se podrían considerar en riesgo por la automatización en 2020, un porcentaje que se prevé que podría aumentar al 21 % a mediados de esta década y al 34 % a partir de 2030. Un informe reciente de Hatzius *et al.* (2023) respalda esta tendencia y estima que la IA generativa podría eliminar unos 300 millones de puestos de trabajo a tiempo completo. Aunque ello podría suponer un aumento del 7 % del PIB mundial, nuevos puestos de trabajo y un auge de la productividad, también podría devaluar algunos empleos y provocar una mayor desigualdad económica, según advierte Greenhouse (2023).

En el ámbito de los medios de comunicación, esta tecnología puede facilitar el trabajo de los profesionales del sector mediático, al ofrecerles herramientas que les permitan generar contenido de forma rápida y eficiente. Sin embargo, también puede entrañar riesgos para la originalidad, la calidad y la veracidad de los contenidos, al facilitar la manipulación y la falsificación. Asimismo, implica nuevos retos y oportunidades para los creadores y los consumidores de contenidos, que deben adquirir nuevas competencias y habilidades, así como una mayor conciencia ética y social sobre el uso responsable de la IA (Davenport y Mittal, 2022).

La creación automatizada de contenido ofrece diversas aplicaciones potenciales. Sin apenas intervención humana, la IA generativa puede producir noticias, reportajes, guiones, traducciones y resúmenes, y permite también recrear la imagen y la voz de personalidades públicas. Estas aplicaciones pueden mejorar la productividad, la accesibilidad y la diversidad de los contenidos mediáticos, pero el contenido automático

también plantea importantes desafíos éticos, legales y sociales relativos a la veracidad, la autoría, la transparencia, la responsabilidad o el impacto en la opinión pública. Los recientes avances en IA presentan, en consecuencia, un panorama complejo del que surgen oportunidades, pero también ciertas limitaciones y amenazas que conviene observar.

Dada la magnitud de los progresos orientados a la producción automatizada de contenidos, hoy es oportuno examinar estos nuevos desarrollos y los efectos que pueden ocasionar en el público receptor. En este trabajo se analiza el estado de la cuestión de los modelos generativos aplicados a los contenidos mediáticos, así como las ventajas y los riesgos que conllevan para el periodismo escrito y audiovisual, para el entretenimiento informativo y, de hecho, para la comunicación social en general, que suele implicar un grado variable de creación mediática, que a su vez queda plasmada en formas de creación como la publicidad, el cine y las series, los documentales, los reportajes o los videoclips.

2. Metodología

Este artículo presenta una revisión exploratoria de diversas fuentes relevantes que abordan el uso de la IA generativa en los medios de comunicación y una reflexión sobre sus implicaciones éticas y legales. Propone así una visión panorámica sobre las ventajas prácticas de este desarrollo tecnológico y sobre la problemática ética, legal y laboral que suscita, así como también sobre su impacto presente y potencial en la sociedad.

La revisión bibliográfica se ha basado en una búsqueda sistemática en las bases de datos bibliográficas *Scopus*, *Web of Science* y *Google Académico*, y en los motores de búsqueda *Bing* y *Google*, empleando una combinación de términos representativos del tema de estudio. Se aplicaron criterios de inclusión y exclusión para filtrar los resultados más pertinentes, tales como la fecha de publicación (preferentemente desde 2020), el idioma (español o inglés), el tipo de documento (artículos periodísticos, técnicos o científicos, y monografías especializadas) y la relevancia temática (trabajos que tratan aspectos éticos y legales de los algoritmos creativos). Se incluyeron además algunos otros documentos que se consideraron relevantes en particular, hallados mediante la aplicación de curaduría algorítmica *Google Discover*. A partir del material escogido se ha elaborado una síntesis integradora y se han identificado las principales tendencias, controversias y desafíos que plantea el uso de la IA generativa en los medios de comunicación desde una perspectiva ética y legal.

3. Un nuevo paradigma en la creación de contenidos

La IA generativa se ha convertido en una herramienta valiosa para la producción de contenido en los medios de comunicación. Esta tecnología ha demostrado ser útil en ámbitos como el periodismo escrito, la creación gráfica, la producción audiovisual y el entretenimiento, entre otros. Con ella se pueden generar textos, voces e imágenes (fijas y en movimiento) de forma cada vez más automatizada, lo que simplifica tareas, reduce tiempos de producción y favorece así la eficiencia.

Pero estos avances conllevan también importantes retos éticos. Entre ellos se encuentran la transparencia y la responsabilidad en el uso de la 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 genere discriminación o acentúe desigualdades. Son cuestiones clave que se deben abordar para asegurar que la tecnología se utilice de manera responsable y se respeten los valores fundamentales de la sociedad.

3.1. Redacción automatizada de noticias

Una de las aplicaciones más prometedoras del aprendizaje profundo es la producción de textos que simulan la redacción humana. Con este objetivo se han elaborado diversos modelos de lenguaje, y el más potente hasta la actualidad es el llamado Generative Pre-trained Transformer, más conocido por las siglas GPT. Creado por el laboratorio OpenAI, este modelo está diseñado para producir textos que imitan de manera convincente el estilo, el tono y la arquitectura de la escritura humana. Y tal es la calidad de los textos que así se generan que resulta difícil distinguirlos de los redactados por una persona.

Las capacidades de GPT no se limitan a la producción de textos, sino que también abarcan la posibilidad de traducirlos, adaptarlos y resumirlos basándose en el estudio del contexto, lo que le confiere a este modelo un gran potencial y numerosas aplicaciones prácticas. Ya antes del lanzamiento de *ChatGPT*, en noviembre de 2022, muchas empresas habían comercializado herramientas apoyadas en GPT, tales como *Anyword*, *Copy.ai*, *Jasper* o *Wordtune*, para la elaboración automática de textos. Hoy existen aplicaciones específicas dirigidas a campos como los 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 periodístico, hace tiempo que se utilizan modelos lingüísticos para convertir datos en noticias narradas, de forma prácticamente autónoma, casi sin intervención humana. La producción automatizada de noticias, que comenzó como un experimento a pequeña escala, se está convirtiendo en un fenómeno global. Sin embargo, el avance es lento. Aunque los medios hablan mucho de esta tecnología, la aplican con timidez en sus procesos por desconfianza y por miedo a la destrucción de empleos (Sánchez García *et al.*, 2023). En cualquier caso, la IA permite hoy producir noticias sobre economía, deportes y meteorología, entre otros temas, y puede complementar la labor de los redactores, por ejemplo, sugiriendo temas de interés basados en la identificación de tendencias.

Algunos medios ya han publicado demostraciones del potencial de esta tecnología. La más notoria llegó de la mano del periódico británico *The Guardian* (OpenAI, 2020). A partir de una premisa para una nota de opinión, el modelo produjo ocho propuestas que los redactores luego combinaron tal como se editaría un artículo cualquiera —recortando y reordenando frases—, solo que dedicando menos tiempo. Aquel experimento, que otros medios han replicado, da una idea de cuál podría ser el futuro del periodismo escrito. Ese futuro, de hecho, ya está aquí, pues un buen número de periódicos (*Daily Express*, *Daily Mirror*, *Wall Street Journal*, etc.) y agencias de noticias (Agence France-Presse, Associated Press, Norsk Telegrambyrå, Reuters, etc.) están usando la generación de lenguaje natural para automatizar la redacción de artículos. 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), esto es, artículos periodísticos organizados en forma de lista y escritos en párrafos cortos, un formato frecuente en Internet por su fácil lectura (Imagen 1).

Imagen 1. Un listículo creado por IA y publicado por el grupo editorial Reach.

From a Victorian market to twitching: 7 things to do with visitors to show off Newport

We've got the answer for the next time you ask yourself, 'Where should we take visitors in Newport?'

 © 2 March 2023



MARKET TREATS: Newport Market is now a hot destination for locals and visitors to the city. Pictured are Ceri Johnston Bower and Chelsea Frampton of Friends in Knead (hal) bakery (Image: WalesOnline)

Stuck for suggestions of things to do with visitors who are coming to stay with you in Newport? Can't, for the life of you, think of an idea even though you've lived in the city longer than the Transporter Bridge has been open?!

Fuente. InYourArea.co.uk (2023).

Esta automatización presenta una ventaja evidente: en teoría, puede dejar más 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 tienden a ser textos asépticos, impersonales y algo superficiales en los que se echa en falta la puesta en contexto, la narrativa y el comentario de una persona experta en el ámbito (Longoni *et al.*, 2022). Por tanto, la escritura automática es una práctica que, si llega a generalizarse, podría propiciar una degradación de los contenidos, lo que empobrecería, en consecuencia, la dieta informativa del público.

Al mismo tiempo, hay dudas más que razonables sobre el valor lícito y legal 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 volúmenes de información que está accesible de forma pública. En la Unión Europea, el *Reglamento General de Protección de Datos* no establece que ese método sea ilegal, si bien restringe lo que se puede hacer con los datos extraídos. La minería de textos está avalada legalmente cuando tiene una finalidad de investigación científica no comercial.

Ciertamente, el uso de la información recogida puede suponer un problema cuando reemplaza 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. Ante esta situación, la Asociación de Medios de Información ha formulado la exigencia de una nueva tasa por el aprovechamiento que la IA hace de sus contenidos sin reconocimiento ni retribución (Castillo, 2023).

También inquietan las posibilidades que la IA abre a la proliferación de contenido engañoso, falso o de mala calidad. Y es un temor fundado, como lo demuestra la eclosión de una nueva ola de granjas de contenido. En agosto de 2023 la organización Newsguard, que evalúa la fiabilidad de las fuentes de información en Internet, encontró 452 sitios de noticias creados casi enteramente por modelos de lenguaje. Se trata de webs de baja calidad que producen muchos artículos con titulares sensacionalistas (*clickbait*), saturados de anuncios para maximizar los ingresos publicitarios. Imitan el aspecto y el lenguaje de los medios tradicionales, pero sin un editor al frente, y los hay que difunden cientos de artículos al día, algunos con narrativas falsas, información errónea o contenido plagiado (Newsguard, 2023). Además, casi todo el contenido presenta un lenguaje superficial y frases repetidas, rasgos típicos del texto generado por IA.

3.2. 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 breves indicaciones expresadas con palabras en lenguaje natural. Para ello, emplean redes generativas antagónicas, esto es, sistemas de IA que se entrenan mediante el aprendizaje profundo (*deep learning*) a partir de grandes cantidades de datos. Este logro tecnológico tiene un lado amable y hasta lúdico, dado que permite producir con suma facilidad imágenes de un realismo sorprendente, pero también esconde problemas y riesgos que requieren atención.

Por una parte, al entrenarse con datos sin filtrar, los modelos generativos pueden reproducir sesgos raciales, culturales y de género (Hao, 2021), lo que implica que los algoritmos no son neutrales y que conviene corregir sus desviaciones para garantizar el principio de justicia. Por otra parte, la confianza en las evidencias gráficas se debilita cuando algunas cabeceras usan imágenes falsas, pero muy realistas, para ilustrar sus titulares, lo que puede propiciar nuevas formas de engaño. En efecto, el contenido gráfico falso puede parecer auténtico, como así lo han demostrado algunas imágenes sintéticas que se viralizaron entre marzo y abril de 2023, tras una importante actualización de *Midjourney*. Estas imágenes mostraban, con un realismo asombroso, escenas falsas como el expresidente Donald Trump forcejeando con la policía (Imagen 2), el Papa Francisco con un abrigo de plumas o un abrazo entre los políticos españoles Yolanda Díaz y Pablo Iglesias. Estas figuraciones, al dar la ilusoria impresión de ser auténticas, pueden propiciar nuevas formas de desinformación.

Desde otro punto de vista, la IA generativa puede tener valor periodístico porque permite crear ilustraciones abstractas y conceptuales. Es una tecnología al alcance de cualquiera, sin necesidad siquiera de formación en diseño gráfico, lo que también entraña un problema. Incluso si el contenido generado no vulnera la imagen

de nadie, sí puede perjudicar a profesionales de la fotografía y la ilustración, que pueden ver desplazada y devaluada su habilidad para presentar trabajos personalizados y de buena calidad (Barandy, 2022).

Imagen 2. Falsas imágenes de Donald Trump arrestado por la policía, generadas con *Midjourney*



Fuente. Higgins (2023).

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. Le siguió la revista *Cosmopolitan*, con una nueva demostración de las capacidades de esta tecnología, usando *DALL·E 2*. Y quizá el caso más polémico fue el protagonizado por el escritor Charlie Warzel, colaborador de la revista *The Atlantic*, que desató una ola de críticas por ilustrar un artículo de opinión con una imagen sintética (Imagen 3). 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 (Warzel, 2022). 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.

La continua aparición de nuevas y mejores herramientas de IA para producir contenido genera inquietud en el sector cultural. Una señal reciente es la huelga del Sindicato de Guionistas de Estados Unidos, que exige, entre otras demandas, que la IA no sustituya labores creativas, ni escriba o reescriba material literario, ni se entrene con obras de guionistas (Broderick, 2023). Además, el 36 % de los trabajadores de la industria estadounidense del entretenimiento temen el impacto de la IA generativa en sus empleos, sobre todo por la vulneración de la propiedad intelectual (Schomer, 2023).

3.3. Vídeos hipertrucados: tecnología *deepfake*

La manipulación de vídeos es una práctica casi tan antigua como el propio contenido audiovisual. Desde sus inicios se han usado técnicas para modificar vídeos, como recortarlos, editarlos o ralentizarlos, con el fin de cambiar su contexto, añadir elementos nuevos, antes inexistentes o, por ejemplo, simular estados de ebriedad de quienes aparecen en ellos. Estas técnicas modifican el material original, pero no la apariencia o las declaraciones de los protagonistas.

Imagen 3. Artículo de Charlie Warzel encabezado por una ilustración artificial



Fuente. *The Atlantic* (2022).

No obstante, en los últimos años ha surgido una nueva forma de crear contenidos falsos que supone un reto ético y social: la tecnología *deepfake*. Esta tecnología emplea la IA generativa para crear vídeos con un nivel de realismo tan alto que cuesta detectar su falsedad. A diferencia de las técnicas anteriores, los vídeos *deepfake* pueden transformar por completo la apariencia o el discurso de los protagonistas, haciéndoles decir o hacer cosas que no son reales (Botha y Pieterse, 2020).

La expresión *deepfake* proviene de la combinación de los términos *deep learning* y *fake* para referirse a un contenido falso generado con técnicas de aprendizaje profundo. Este contenido, que distorsiona la realidad para hacer creer algo irreal, se conoce también como *ultrafalso* o *hipertrucado*, dos palabras que reflejan bien la naturaleza de estas imágenes, falseadas de forma tan radical, tal es la sofisticación de esta tecnología, que apenas se percibe el engaño.

La recreación realista de personas tuvo su origen en el cine mediante la técnica de CGI (*computer-generated imagery*). Esta tecnología facilitó la sustitución digital del rostro de artistas fallecidos durante el rodaje por el de un doble, como se hizo, por ejemplo, con Oliver Reed en *Gladiator* (2000) o con Paul Walker en *Fast & Furious 7* (2015). El séptimo arte también innovó con el llamado «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 empleado después por marcas como Peugeot, Mercedes Benz o Virgin Trains en sus campañas publicitarias (Imagen 4). No obstante, estas técnicas tienen el inconveniente de resultar costosas, ya que exigen contar con personal especializado que dedique mucho tiempo al proceso.

Imagen 4. Protagonistas de *Con la muerte en los talones* (1959) insertados en un tren de Virgin Trains (2005), ante la sorpresa de una usuaria moderna



Fuente. AdForum.

La IA generativa ha supuesto un gran avance, y se ha vuelto tan accesible que el CGI parece hoy anticuado. La tecnología *deepfake* ha revolucionado la recreación de personas en las industrias creativas. Entre sus aplicaciones destacables se encuentran la producción de material para el cine o 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» con fines publicitarios 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». Aquí, el presentador Gran Wyoming y su colaborador Dani Mateo encarnan a figuras políticas reconocidas. Esta misma aplicación de la IA ha impulsado, en el entorno de Internet, la creación y difusión de vídeos más o menos divertidos en los que el rostro de ciertas estrellas, como Steve Buscemi o Nicolas Cage, suplantando a personajes de cualquier película. Y también ha dado una enorme popularidad a canales de *TikTok* en los que varios imitadores suplantando en clave de parodia a celebridades como Tom Cruise, Keanu Reeves o Margot Robbie.

La falsificación de contenido audiovisual no es un fenómeno nuevo. La novedad es el grado de simplificación que la IA ha aportado al proceso. Antes se requería 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 lo ficticio y lo real se confunden.

Esto ha generado controversia en el ámbito del entretenimiento informativo. En mayo de 2022, el canal de televisión France 3 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. Entre las figuras invitadas se encuentran Lady Di, François Mitterrand y Dalida, la diva de la canción francesa a quien, en un momento televisivo muy extraño, el presentador le pregunta nada menos que por su suicidio (Imagen 5). 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?

Además de la falta de consentimiento de las personas recreadas por IA, que por ahora se solventa con el permiso de los familiares, 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 nos llevan más allá de ese valle, a un mundo donde el engaño es más sutil y donde

las caras producidas por IA no solo se confunden con las reales, sino que además generan más confianza (Nightingale y Farid, 2022).

Imagen 5. Un instante del programa *Hôtel du temps*.



Fuente. France 3 (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 e insospechadas 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). Los algoritmos de aprendizaje profundo se entrenan con la infinidad de imágenes que brinda Internet, pero la tecnología avanza tan rápido que cada vez necesitan menos datos de entrada para lograr un nivel similar de realismo (Giansiracusa, 2021).

3.4. Síntesis de vídeo a partir de texto

Los modelos generativos también permiten crear vídeos sintéticos a partir de simples indicaciones de texto. Esta tecnología, que se está desarrollando sobre la base de la síntesis de imágenes, está aún en sus inicios, pero ya evidencia un gran potencial creativo, como así lo demuestran desarrollos como *Make-A-Video*, de Meta, *NUWA-XL*, de Microsoft, y *Phenaki*, de Google.

Con estas herramientas se pueden crear secuencias de vídeo en cuestión de minutos, sin requerir material audiovisual previo, lo que puede suponer una valiosa herramienta para artistas. Solo se necesita facilitar una base de texto sobre la que trabajar y el *software* genera un vídeo listo para usar.

A diferencia del realismo que ya se obtiene con algunas herramientas de síntesis gráfica, los resultados que ofrece la síntesis de vídeo distan aún de ser perfectos. Es probable que esta tecnología evolucione a un ritmo más lento que la síntesis de imágenes, dado que es más complejo generar vídeos que fotos; además de generar correctamente cada uno de los píxeles, el sistema también ha de predecir cómo evolucionarán en una secuencia temporal.

Aun así, las mejoras que cabe esperar en esta técnica también plantean importantes riesgos éticos y sociales que convendrá vigilar de cerca. La capacidad de crear vídeos sintéticos cada vez más realistas puede

dar lugar a usos malintencionados, como los que ya propicia la tecnología *deepfake*, según ha quedado demostrado.

En el ámbito de las narrativas audiovisuales, la síntesis de vídeo con IA generativa no muestra aún una gran innovación u originalidad, sino que se limita a reproducir los estilos visuales de épocas anteriores del cine y la televisión. Según señala López Delacruz (2023), los vídeos sintéticos que circulan en plataformas de consumo audiovisual se inscriben en una estética retro y nostálgica que evoca un pasado cultural sin una finalidad crítica o reflexiva. Así, estas narrativas renuncian al potencial novedoso y creativo de la IA generativa, y se conforman con la imitación de expresiones culturales precedentes como forma de consumo digital. Pero la falta de originalidad no es exclusiva de esta nueva forma de creación, pues también aqueja a las plataformas de *streaming* desde hace tiempo (Solà, 2021).

3.5. Clonación de voz

La síntesis del habla es una tecnología sencilla que permite convertir texto en una voz muy parecida a la humana, y que se emplea 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, los avances en el aprendizaje profundo han 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. Y, claro, 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 sin por ello presentar las limitaciones y las lógicas exigencias humanas. No obstante, el sector del doblaje teme por su futuro, ante las prometedoras posibilidades de la síntesis de voz (Pérez Colomé, 2023). Se están haciendo progresos en esta tecnología para doblar películas y series artificialmente imitando el timbre y el tono de una voz original.

Los medios de comunicación, por su parte, pueden usar voces clonadas para otros fines. Por ejemplo, para narrar noticias y eventos con la voz de un periodista o presentador famoso y hacer así más atractiva la información. O para dar voz a personajes históricos o fallecidos en documentales y programas de televisión cuando ya no se puede grabar su propia voz.

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 y escritor estadounidense, que se había suicidado poco antes (Imagen 6). Para lograr que Bourdain pronunciase 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 señal de identidad, y falsificarla en un documental, que es una expresión de la realidad, puede provocar que el público se sienta engañado. Un documental exige honestidad en la delimitación de los elementos reales y ficticios. Si no se explicita la naturaleza del discurso desde el inicio, se 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.

Imagen 6. Un fotograma del documental *Anthony Bourdain: un chef por el mundo* (2021).



Fuente. IMDb.

4. Discusión y conclusiones

La IA generativa se ha convertido en una tecnología de alto impacto para el periodismo y los medios de comunicación y, por lo tanto, de creciente relevancia en la producción de contenidos. Con su irrupción en las tareas creativas, los ordenadores han dejado de ser herramientas de apoyo para pasar a ser protagonistas, capaces de crear contenido original. A partir de unas indicaciones sencillas, una descripción simple o unos parámetros básicos, un algoritmo es capaz de generar contenido nuevo —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 y ubicuo. 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, bases de datos científicas, y programas de edición de imagen y vídeo.

Sin embargo, esta tecnología también conlleva importantes riesgos éticos y sociales; a medida que avanza, aumentan las posibilidades de engaño y otros peligros. Los contenidos artificiales son tan verosímiles que pueden confundir a la sociedad si no hay un código de conducta. Si muchas personas ya tienen dificultades para desenmascarar un bulo o un engaño con evidencias simples, el panorama que dibuja esta creciente sofisticación resulta aún más desalentador y preocupante. Por eso, es necesario aumentar la conciencia pública sobre el uso responsable y ético de esta tecnología.

4.1. Implicaciones del contenido artificial

La capacidad de la IA generativa para crear automáticamente todo tipo de contenidos supone una innovación radical en la forma de producir contenido y conlleva implicaciones en la propiedad intelectual, la veracidad informativa, la identidad personal y la creatividad humana. También plantea desafíos éticos en esferas clave de la sociedad, tales como la democracia, el empleo y el sistema económico.

4.1.1. Suplantación de la escritura humana

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 un modelo de lenguaje se pueden falsear noticias sin apenas esfuerzo y de forma convincente, imitando el estilo de las informaciones redactadas por humanos, como ya hace una nueva generación de granjas de contenido (Newsguard, 2023). Esta posibilidad supone una amenaza para la credibilidad del periodismo y para el derecho de la ciudadanía a la información veraz.

Además, los modelos de lenguaje pueden generar contenido tóxico, incluso en contextos benignos (Ousidhoum, 2021). 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). Es ya bastante evidente que la IA complicará, aún más, el acceso a la verdad.

Los gigantes de las búsquedas, por su parte, tampoco pueden hacer mucho para detectar textos automáticos y advertir de ello. Google, por ejemplo, califica de fraudulento el contenido escrito por una IA, y así lo apunta en sus directrices de calidad, que se oponen sin atenuantes al texto generado por procesos automáticos. Sin embargo, y esto es paradójico, no cuenta con medios para identificarlo y, por lo tanto, no puede señalarlo ni tampoco penalizarlo.

Por otro lado, aunque los modelos de lenguaje son versátiles y producen textos naturales, con una fluidez narrativa y de conversación sorprendente, también tienen limitaciones. Una de ellas es que no cubren todo el corpus de información accesible en Internet. Solo abarcan una selección de datos relevantes, considerados de calidad, que en su mayoría proceden de fuentes de habla inglesa, lo que además compromete la diversidad cultural (Navigli, Conia y Ross, 2023). Aun así, se están haciendo esfuerzos para subsanar esa limitación complementando las respuestas con el resultado de búsquedas en tiempo real.

Otra limitación de los modelos de lenguaje es que no siempre se apoyan en fuentes fiables o evidencias robustas, que son lo que determina la veracidad o la solidez de un argumento. Los modelos de lenguaje solo se limitan a decir, con una coherencia solo aparente, cómo debería sonar una respuesta. Por lo tanto, el rigor no está garantizado, y el contenido generado, que no es estrictamente información, sino una mera construcción sintética, se debe contrastar con fuentes fiables y personas expertas en la materia.

La IA generativa es válida entonces para situaciones que admiten cierto margen de error, cierta superficialidad argumental, incluso algún disparate. Pero no lo es para cuestiones críticas, como un trabajo científico, un consejo legal o financiero, o una consulta médica. Produce una engañosa ilusión de pensamiento racional, pero no razona ni dispone de conocimiento fiable sobre el mundo. No entiende, en un sentido humano, nada de lo que escribe. El texto que produce no es necesariamente veraz. Aunque suene elocuente y hasta convincente, el contenido puede ser incorrecto o absurdo, porque la IA, a veces, se inventa hechos, personas, datos o fuentes. Puede enredarnos con el aplomo de un mentiroso compulsivo. Y ni siquiera se puede afirmar que mienta, ya que ni siquiera dispone de un modelo de verdad.

4.1.2. Suplantación de la creación gráfica

Preocupa también la práctica cada vez más frecuente entre los medios digitales de ilustrar sus noticias con imágenes sintéticas. Esta tendencia socava la confianza que se tiene en las evidencias gráficas y aumenta la confusión. Y 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, las buenas prácticas en este aspecto, y que los agregadores de noticias (*Google Noticias*, *Flipboard*, *Yahoo Noticias*, etc.) las 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 los temores 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, dado que 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 propio trabajo. De hecho, ya parece evidente que los ilustradores podrían estar entre los primeros artistas en ser desplazados por la IA (Barandy, 2022). Ante semejante 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 admisible o razonable de 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.

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 que ya estaban 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 no basta con que un sistema sea competente, sino que además debe actuar con ética. El sistema se debe calibrar y reentrenar para que evite perpetuar sesgos perjudiciales y mejore así la calidad de los resultados.

4.1.3. Suplantación de la realidad: vídeos sintéticos y ultrafalsos

En el terreno 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 estas posibilidades, que están cada vez más al alcance de la mano, radica en que el realismo de estos vídeos ultrafalsos los hace muy creíbles a la vez que difíciles de verificar. Aunque falsos, parecen de verdad, tienen aspecto de clara evidencia. Por tanto, tienen un gran poder para manipular la opinión pública. Entonces, suponen una amenaza para la imagen de las personas, desde el ciudadano anónimo hasta un político o una celebridad, sobre todo porque la tecnología *deepfake* está hoy 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. Una de las propuestas para evitar el engaño de los contenidos artificiales consistiría en 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 mundo 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 famosa bailaora. 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?

4.1.4. Suplantación de la voz humana

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 seña 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 normas éticas y han incorporado medios para asegurar que la persona dueña de la voz sintetizada realmente ha dado su consentimiento. Con estas medidas buscan proteger los intereses de quienes prestan su voz, para que puedan controlar su uso y recibir unas regalías justas.

Asimismo, se debe informar al público cuando se utilice una voz clonada, para evitar confusiones o engaños. Es una forma de respetar el derecho a la información y a la transparencia, así como de proteger la dignidad y la reputación de las personas cuya voz se simula artificialmente. El sector del doblaje reclama que las voces sintéticas tengan un «acento IA» que las distinga al oír las, aunque ello no impedirá que esta tecnología desplace a los estudios de doblaje (Pérez Colomé, 2023).

También se debe tener en cuenta el impacto emocional y psicológico que puede tener escuchar una voz clonada, sobre todo si recrea la de una persona fallecida o ausente. Se debe evitar el uso imprudente o manipulador de las voces clonadas, dado que puede molestar o causar sufrimiento a las personas que las oyen o a sus familiares y allegados.

4.2. Hacia una ética de la IA

Son muchos los retos que plantea este gran logro tecnológico. La IA generativa puede mejorar la productividad, pero puede también causar una enorme disrupción. Este cambio de paradigma plantea dilemas éticos en áreas como la democracia, el empleo, el sistema económico y la cultura. Afrontar estos desafíos exige un enfoque interdisciplinar y participativo que apueste por reforzar la educación, la ciencia y la democracia. Las empresas tecnológicas tienen cada vez más poder, y eso es un problema para la democracia. Por eso, es urgente regular los algoritmos y a sus responsables, así como promover una visión crítica y consciente de la tecnología entre la población.

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 velar tanto por el acceso como por 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. 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 (Metz, 2022).

Hay cierta inclinación a conceder cualidades humanas a la IA y a atribuirle 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, es innegable que esta tecnología ofrece una valiosa ayuda para agilizar una buena cantidad de labores en las que el ser humano es menos eficiente.

Los algoritmos creativos abren un horizonte de innovación en muchos sectores y actividades, pero también plantean inquietudes éticas sobre su uso y regulación. La IA generativa cambiará el paradigma de la creación de contenidos, que ya no dependerá solo del talento humano. Se avecinan, pues, tiempos de transformación y nuevas formas de crear, con reglas de juego distintas. Las posibilidades son prometedoras y despiertan curiosidad e interés, pero también miedo, recelo y preocupación.

Asimismo, 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 mitigar o compensar posibles efectos negativos.

El Parlamento Europeo ya ha sentado 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

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.

La actual regulación de la IA supone un primer paso hacia una gobernanza integral de esta tecnología. El reglamento europeo establece criterios para clasificar la IA y exigir transparencia según el nivel de riesgo que implique. Sin embargo, esta normativa debe ajustarse a la realidad de los usos de la IA, lo que requiere una colaboración entre el sector público y el privado. Un ejemplo de esta colaboración es el banco de pruebas impulsado por el Gobierno de España y la Comisión Europea, que busca definir buenas prácticas para la implementación de la regulación europea de la IA. España ha creado además la Agencia Española de Supervisión de la Inteligencia Artificial, siendo así el primer país europeo en dotarse de una autoridad supervisora, que será clave para favorecer la correcta interpretación y el cumplimiento de la ley.

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 programas de concienciación para que el público comprenda el alcance de la IA generativa y aprenda a defenderse de las nuevas falsificaciones. La educación, ciertamente, juega un papel clave para preparar a la ciudadanía ante la proliferación de contenidos artificiales, tanto en lo que se refiere al aprendizaje sobre la IA como a la capacidad de la sociedad para adaptarse a sus efectos. La UNESCO aboga por un enfoque que fomente la inclusión, la equidad y el respeto a los derechos humanos, así como el desarrollo de competencias básicas para esta nueva era (Giannini, 2023). Entre estas competencias, será esencial estimular el pensamiento crítico para evaluar la calidad, la veracidad y la relevancia de los contenidos generados por IA.

4.3. Futuras líneas de investigación

Según ha quedado de manifiesto, la IA generativa presenta aplicaciones en diversos campos y sectores, pero también plantea importantes retos y dilemas para su regulación y control, ya que puede generar contenido falso o engañoso que afecte a los derechos y valores de las personas y la sociedad. Por ello, es necesario investigar más sobre esta tecnología, sus beneficios y sus riesgos, desde una perspectiva multidisciplinar y participativa. Algunas de las posibles líneas de investigación futuras sobre los ámbitos de la IA generativa y la regulación legal que necesita podrían ser las siguientes:

- Ampliar el estudio del impacto social, económico y cultural de la IA generativa en diferentes sectores y ámbitos, como la educación, la salud, el arte, el periodismo o el entretenimiento.
- Analizar los desafíos técnicos y metodológicos para garantizar la calidad, la fiabilidad y la seguridad de los sistemas de IA generativa, así como para evaluar su rendimiento y su valor añadido.
- Desarrollar herramientas y estrategias para detectar, prevenir y combatir las falsificaciones generadas por IA, como los *deepfakes* o los textos falsos, que puedan afectar a la veracidad de la información, a la reputación de una persona, un colectivo o una institución, o incluso a la libertad de elegir en democracia.
- Explorar las implicaciones éticas y morales de la creación y el uso de contenido generado por IA, especialmente en relación con los derechos de autor, la propiedad intelectual y la responsabilidad legal.
- Comparar y contrastar los diferentes marcos normativos y regulatorios existentes o propuestos sobre la IA generativa a escala nacional, regional e internacional, identificando sus fortalezas, debilidades y oportunidades de mejora.
- Investigar las percepciones, actitudes y comportamientos de los usuarios y los consumidores de contenido artificial, así como sus necesidades, expectativas y preferencias.
- Diseñar e implementar programas de formación y sensibilización sobre la IA generativa dirigidos a diferentes públicos y agentes sociales, como profesionales, educadores, periodistas o ciudadanos.

Referencias

- Ajder, H. (2020). *Deepfake threat intelligence: a statistics snapshot from June 2020*. Sensity.
- Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). *The state of deepfakes: landscape, threats, and impact*. DeepTrace. <https://is.gd/LK0X5N>
- Ayuso, S., y Pascual, M. G. (2023, 11 de mayo). Europa quiere poner más obligaciones a la inteligencia artificial generativa como la de ChatGPT. *El País*. <https://is.gd/IQNRIA>
- Barandy, K. (2022, 10 de agosto). Will artists be replaced by artificial intelligence? *Designboom*. <https://is.gd/NZqVHX>
- Bender, E. M. (2022, 14 de junio). Human-like programs abuse our empathy: even Google engineers aren't immune. *The Guardian*. <https://is.gd/rAUdbF>
- Bhargava, C., y Sharma, P.K. (Eds.) (2022). *Artificial intelligence: fundamentals and applications*. CRC Press.
- Boden, M. A. (2018). *Artificial intelligence: a very short introduction*. Oxford University Press.
- Botha, J., y Pieterse, H. (2020). Fake news and deepfakes: a dangerous threat for 21st century information security. *15th International Conference on Cyber Warfare and Security: ICCWS 2020*, 57-66.
- Boucher, P. (2020). *Artificial intelligence: how does it work, why does it matter, and what can we do about it?* Servicio de Estudios del Parlamento Europeo. <https://doi.org/10.2861/44572>
- Broderick, R. (2023, 31 de mayo). AI can't replace humans yet: but if the WGA writers don't win, it might not matter. *Polygon*. <https://is.gd/PT9hSr>
- Campesato, O. (2020). *Artificial intelligence, machine learning and deep learning*. Mercury Learning and Information.
- Castillo, C. (2023, 3 de mayo). Los creadores del canon AEDE quieren una "tasa ChatGPT" para la inteligencia artificial. *elDiario.es*. https://eldiario.es/1_9b351c
- Dean, I. (2022, 11 de agosto). You can now sell your DALL·E 2 art, but it feels murky. *Creative bloq*. <https://is.gd/ov3ln5>
- Dale, R. (2022). The voice synthesis business: 2022 update. *Natural language engineering*, 28(3), 40-408. <https://doi.org/10.1017/S1351324922000146>
- Davenport, T. H., y Mittal, N. (2022, 14 de noviembre). How generative AI is changing creative work. *Harvard Business Review*. <https://is.gd/by7hQt>
- Giannini, S. (2023). *Generative AI and the future of education*. UNESCO. <https://is.gd/CbhGO5>
- Giansiracusa, N. (2021). *How algorithms create and prevent fake news: exploring the impacts of social media, deepfakes, GPT-3 and more*. Apress. <https://doi.org/10.1007/978-1-4842-7155-1>
- Greenhouse, (2023, 8 de febrero). US experts warn AI likely to kill off jobs and widen wealth inequality. *The Guardian*. <https://is.gd/n38xQn>
- Hao, K. (2021, 3 de febrero). Internet está tan sesgado que, para la IA, las mujeres solo llevan bikini. *MIT Technology Review*. <https://is.gd/kSOd56>
- Hatzius, J., Briggs, J., Kodnani, D., y Pierdomenico, G. (2023, 26 de marzo). *The potentially large effects of artificial intelligence on economic growth*. Goldman Sachs.
- Higgins, E. (2023, 20 de marzo). Making pictures of Trump getting arrested while waiting for Trump's arrest. *Twitter*. <https://is.gd/oiAwPp>
- Kreps, S., McCain, R. M., y Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, 9(1), 104-117. <https://doi.org/10.1017/xps.2020.37>
- Longoni, C., Fradkin, A., Cian, L., y Pennycook, G. (2022, junio). News from generative artificial intelligence is believed less. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 97-106. <https://doi.org/10.1145/3531146.3533077>
- López Delacruz, S. (2023). Un vínculo paradójico: narrativas audiovisuales generadas por inteligencia artificial, entre el pastiche y la cancelación del futuro. *Hipertext.net*, 26, 31-35. <https://doi.org/10.31009/hipertext.net.2023.i26.05>
- Lyons, B. A., Montgomery, J. M., Guess, A. M., Nyhan, B., y Reifler, J. (2021). Overconfidence in news judgments is associated with false news susceptibility. *PNAS*, 118(23), e2019527118. <https://doi.org/10.1073/pnas.2019527118>
- Metz, C. (2022, 5 de agosto). AI is not sentient: why do people say it is? *The New York Times*. <https://is.gd/gBlLu2>
- Navigli, R., Conia, S., y Ross, B. (2023). Biases in large language models: origins, inventory and discussion. *Journal of Data and Information Quality*, 15(2). <https://doi.org/10.1145/3597307>
- Nightingale, S. J., y Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *PNAS*, 119(8), e2120481119. <https://doi.org/10.1073/pnas.2120481119>
- Newsguard (2023). *Reports about online misinformation and disinformation from NewsGuard's analysts*. <https://newsguardtech.com/reports>

- OpenAI (2020, 8 de septiembre). A robot wrote this entire article: are you scared yet, human? *The Guardian*. <https://is.gd/CLK6NQ>
- Osmanovic-Thunström, A. (2022, 30 de junio). We asked GPT-3 to write an academic paper about itself, then we tried to get it published. *Scientific American*. <https://is.gd/OnGPRf>
- Ousidhoum, N., Zhao, X., Fang, T., Song, Y., y Yeung, D. Y. (2021). Probing toxic content in large pre-trained language models. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 1*, <https://doi.org/10.18653/v1/2021.acl-long.329>
- Pérez Colomé, J. (2023, 11 de abril). Los actores de voz se unen al ver peligrar su trabajo por la inteligencia artificial. *El País*. <https://is.gd/UzwnED>
- PwC (2018). *Will robots really steal our jobs? An international analysis of the potential long-term impact of automation*. <https://is.gd/ApKuPI>
- Sánchez-García, P., Merayo-Álvarez, N., Calvo-Barbero, C., y Díez-Gracia, A. (2023). Desarrollo tecnológico español de la inteligencia artificial aplicada al periodismo: empresas y herramientas de documentación, producción y distribución de información. *Profesional de la información, 32(2)*, e320208. <https://doi.org/10.3145/epi.2023.mar.08>
- Schomer, A. (2023, 6 de julio). Entertainment industry has high anxiety about generative AI: survey. *Variety*. <https://is.gd/yTnheu>
- Solà, P. (2021, 27 de enero). El síndrome de Lupin y la falta de originalidad de Netflix. *La Vanguardia*. <https://is.gd/rpt5hZ>
- Steed, R., y Caliskan, A. (2021). Image representations learned with unsupervised pre-training contain human-like biases. *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability and Transparency, 701-713*. <https://doi.org/10.1145/3442188.3445932>
- Sweney, M. (2023, 7 de marzo). *Mirror and Express* owner publishes first articles written using AI. *The Guardian*. <https://is.gd/osl6u2>
- Warzel, C. (2022, 17 de agosto). I went viral in the bad way. *The Atlantic*. <https://is.gd/4muxwS>
- Yu, N., Skripniuk, V., Abdelnabi, S., y Fritz, M. (2021). Artificial fingerprinting for generative models: rooting deepfake attribution in training data. *Proceedings of the IEEE/CVF International conference on computer vision, 14.448-14.457*. <https://doi.org/10.1109/iccv48922.2021.01418>

Breve CV del autor

Jorge Franganillo es doctor en Información y Comunicación (Universidad de Barcelona, UB). Ejerce como profesor en la Facultad de Información y Medios Audiovisuales de la UB y ha impartido docencia en universidades de Alemania, Croacia y República Checa. Es miembro del Centro de Investigación en Información, Comunicación y Cultura (CRICC), donde centra sus principales líneas de investigación en la comunicación digital, la documentación audiovisual, la inteligencia artificial generativa y los fenómenos de desinformación. Ha publicado dos libros y numerosos artículos académicos, y colabora como divulgador en la prensa, donde insiste en remarcar la dimensión humana de la tecnología.

Financiación

Este trabajo forma parte del proyecto I+D+i PID2019-107748RB-I00, financiado por el Ministerio de Ciencia e Innovación del Gobierno de España.

Conflicto de intereses

El autor declara la inexistencia de conflicto de intereses.