

# La piratería de la inteligencia artificial: el uso de las obras en el entrenamiento de modelos de IA generativos

## JORGE LUIS ORDELIN FONT

Investigador  
Centro de Investigación y Docencia  
Económicas (CIDE)  
México, México

ORCID: <http://orcid.org/0000-0001-8778-882X>

[jorge.ordelin@cide.edu](mailto:jorge.ordelin@cide.edu)

RECIBIDO: 15 DICIEMBRE 2024.

ACEPTADO: 18 ENERO 2025

## RESUMEN

El objetivo general explicar por qué el uso de las obras protegidas por los derechos de autor en el entrenamiento de los sistemas de inteligencia artificial podría ser considerado como piratería de los derechos de autor. Para su elaboración se utiliza una metodología cualitativa. A partir de tener en cuenta el uso de las obras, la tecnología y los modelos de negocios que se desarrollan alrededor de esta se concluye que sí es posible concebir al uso de las obras en el entrenamiento de los sistemas de IA generativos como un supuesto de piratería, que no solo constituye una infracción de los derechos de autor, sino también, por el impacto que dicho uso tiene en la economía y la creatividad.

**Palabras claves:** Piratería, inteligencia artificial generativa; derechos de autor; obras; datos; entrenamiento de sistemas de IA.

## THE PIRACY OF ARTIFICIAL INTELLIGENCE: THE USE OF ARTWORKS IN THE TRAINING OF GENERATIVE AI MODELS

### ABSTRACT

The general objective is to explain why the use of copyrighted works in the training of artificial intelligence systems could be considered as copyright piracy. A qualitative methodology is used for its elaboration. Considering the use of the artworks, the technology, and the business models that are developed around it, it is concluded that it is possible to conceive the use of the artworks in the training of generative AI systems as a case of piracy, which not only constitutes an infringement of copyright, but also because of the impact that such use has on the economy and creativity.

**Keywords:** Piracy, generative artificial intelligence, copyright, artworks, data, AI system training.

SUMARIO: I. INTRODUCCIÓN. II. PIRATERÍA E INTELIGENCIA ARTIFICIAL. III. OBRAS, TECNOLOGÍAS Y DERECHOS DE AUTOR. 1 La apreciación de limitaciones o excepciones a los derechos de autor. IV. EL IMPACTO ECONÓMICO, LA OTRA CARA DE LA PIRATERÍA. 1. Los beneficios para los desarrolladores de los sistemas de IA generativos. 2. Las pérdidas económicas de los creadores y demás titulares por el uso de sus obras. V. A MANERA DE CONCLUSIONES. VI. REFERENCIAS BIBLIOGRÁFICAS.

CONTENTS: I. INTRODUCTION. II. PIRACY AND ARTIFICIAL INTELLIGENCE. III. ARTWORKS, TECHNOLOGIES, AND COPYRIGHT. 1. The assessment of limitations or exceptions to copyright. IV. THE ECONOMIC IMPACT, THE OTHER SIDE OF PIRACY. 1. The benefits for developers of generative AI systems. 2. The economic losses of creators and other owners for the use of their artworks. V. CONCLUSIONS. VI. BIBLIOGRAPHICAL REFERENCES.

## INTRODUCCIÓN

La demanda colectiva de Sarah Andersen, Kelly McKernan y Karla Ortiz, en nombre propio, y en el de todas las demás personas en situación similar, contra Stability AI Ltd., y Stability AI Inc. Midjourney, Inc. (“Midjourney”); y DeviantArt, Inc. (“DeviantArt”) por infracción directa y indirecta de derechos de autor (en adelante, demanda colectiva)<sup>1</sup>; y la demanda de Getty Images (US) Vs. Stability AI<sup>2</sup>, marcaron el inicio del año 2023 en materia de derechos de autor. Aun cuando son demandas que inician procesos judiciales, de los cuáles no sabemos cuales serán sus resultados, con su interposición se evidencia un fenómeno tecnológico relativamente reciente pero con un importante impacto en las relaciones de derechos de autor en el ámbito digital, el uso de obras como datos de entrenamiento por parte de los sistemas de inteligencia artificial generativos (en adelante sistemas de IA).

Se calcula que Stability AI, por ejemplo, copió más de 12 millones de fotografías de la colección de Getty Images sin su permiso, de conjunto con los subtítulos y metadatos asociados, e igualmente raspó y copió más de cinco mil millones de imágenes de sitios web, por solo mencionar algunas cifras<sup>3</sup>. Los datos reflejados, en principio, comienzan a mostrar una de las preocupaciones más recientes que surgen desde el ámbito de la propiedad intelectual, el uso de grandes cantidades de obras como datos de entrenamiento por los sistemas de IA. Sobre dicho uso surgen innumerables preguntas, en particular, si es legal no contar con su autorización para ello.

La respuesta es compleja y no tiene una única arista de análisis. Sin embargo, si se tiene en cuenta el uso que se hace de las obras, la tecnología y los modelos de negocios que se desenvuelven en relación a esta última, sí debería ser considerada ilegal dicha utilización y, en particular, ser considerada un

supuesto de piratería de los derechos de autor<sup>4</sup>. En correspondencia con ello, el presente artículo se plantea como pregunta de investigación ¿podría ser considerado el uso de las obras protegidas por los derechos de autor en el entrenamiento de sistemas de inteligencia artificial creativos como piratería? Siendo su objetivo general explicar por qué el uso de las obras protegidas por los derechos de autor en el entrenamiento de los sistemas de inteligencia artificial debería ser considerado como piratería de los derechos de autor.

Reconocer que el uso sin autorización de las obras para entrenar sistemas de IA generativos constituye un acto de piratería o, al menos, una infracción de los derechos de autor es importante no solo desde un punto de vista jurídico, sino también económico. Con ello, no solo se abre la posibilidad de establecer demandas jurídicas, sino también, el reconocimiento a los autores de poder participar de los beneficios económicos que se obtienen a partir del desarrollo de modelos de negocios que se despliegan alrededor de estos sistemas de IA.

Sin embargo, debe reconocerse que el estudio del fenómeno de la piratería no es algo sencillo, tiene múltiples aristas y aspectos a tratar, ello no solo lo convierte en complejo sino que también excede, por mucho, los límites de un trabajo de esta naturaleza. Es por ello que este artículo no persigue ofrecer respuestas acabadas sobre el fenómeno, sino plantear dudas e interrogantes que en un corto período deberán ser respondidas por el Derecho. El artículo toma como punto de referencia el hecho de que el fenómeno de la piratería debe abordarse tanto desde la perspectiva del uso de las obras sin autorización, como desde el alcance del desplazamiento, es decir, cuántas obras se habrían comprado y/o distribuido si no se hubieran pirateado, y los ingresos que se hubieran percibidos<sup>5</sup>. Desde esta perspectiva se divide en tres partes. La primera destinada a explicar si es posible que el concepto de piratería pueda ser aplicado al uso de las obras en el entrenamiento de los sistemas de IA; la segunda, aborda la relación que existe entre la tecnología de IA, el uso de datos de entrenamiento y las infracciones de los derechos de autor, incluyendo la posible consideración como excepción y/o limitación de dicha utilización y; por último, los posibles efectos que desde el ámbito de la sustitución provoca dicho uso.

1 SARAH ANDERSEN, an individual; KELLY MCKERNAN, an individual; KARLA ORTIZ, an individual, Individual and Representative Plaintiffs, v. STABILITY AI LTD., a UK corporation; STABILITY AI, INC., a Delaware corporation; MIDJOURNEY, INC., a Delaware corporation; DEVIANTART INC, a Delaware corporation, Class Action Complaint, United States District Court Northern District Of California San Francisco Division, Case 3:23-cv-00201, disponible en <https://stablediffusionlitigation.com/pdf/00201/1-1-stable-diffusion-complaint.pdf>, consultado el 20 de enero de 2023.

2 GETTY IMAGES (US), Inc. v. STABILITY AI, Inc. Demand for Jury Trial, United States District Court for The District of Delaware, Case 1:23-cv-00135-UNA, disponible en <https://www.courtlistener.com/docket/66788385/getty-images-us-inc-v-stability-ai-inc/>, consultado el 15 de febrero de 2023.

3 GETTY IMAGES (US), Inc. v. STABILITY AI, Inc. Demand for Jury Trial, United States District Court for The District of Delaware, Case 1:23-cv-00135-UNA, disponible en <https://www.courtlistener.com/docket/66788385/getty-images-us-inc-v-stability-ai-inc/>, consultado el 15 de febrero de 2023.

4 Los términos de falsificación y piratería se utilizan de manera indistintas en múltiples informes y documentos, aún cuando se pudiera hacer referencia de una forma más específica a los términos de falsificación de marcas y a la piratería de derechos de autor. OFFICE OF THE UNITED STATES TRADE REPRESENTATIVE, (2022A), 2022 Special 301 Report, disponible en <https://ustr.gov/sites/default/files/IssueAreas/IP/2022%20Special%20301%20Report.pdf>, consultado el 24 de febrero de 2023, pág. 9. A los efectos de este trabajo se utilizará el concepto de piratería.

5 BLACKBURN; EISENACH; HARRISON (2019), pág. 2.

Para su elaboración se utiliza una metodología cualitativa, sustentada en métodos deductivos y de análisis teórico-lógicos, los cuales permiten concluir que si se interpreta de manera contextualizada el uso de obras protegidas por los derechos de autor en el entrenamiento de los sistemas de IA generativos estamos en presencia de un uso no autorizado, pirata; no solo porque supone una infracción en materia de los derechos de autor, sino también, por el impacto que dicho uso tiene en la economía y la creatividad, en particular por sus efectos directos e indirectos en esta.

## PIRATERÍA E INTELIGENCIA ARTIFICIAL

Según la Organización Mundial del Comercio (OMC), la piratería no es más que la reproducción no autorizada de materiales protegidos por derechos de propiedad intelectual (como derecho de autor, marcas de fábrica o de comercio, patentes, indicaciones geográficas, etc.) hecha con fines comerciales, así como el comercio no autorizado de los materiales reproducidos<sup>6</sup>. STERLING citado por PANETHIERE, considera que en su acepción habitual no es más que “la actividad consistente en fabricar copias no autorizadas (copias pirata) de obras protegidas y de distribuirlas o venderlas”<sup>7</sup>. Para Freitas, en una concepción similar aunque un poco más amplia, no es más que la reproducción, con miras a obtener un provecho de la propiedad de un titular de derecho de autor, sin su permiso<sup>8</sup>.

El concepto de piratería se ha ido redimensionando tras la aparición de nuevas tecnologías y formas de ejecución. La llegada de Internet supuso que además de los bienes físicos pirateados a partir del comercio electrónico, aparecieran otras formas de piratería, en particular la que tiene lugar en relación con los bienes digitales. En este contexto, se ha señalado que la piratería digital no es más que la copia o distribución ilegal a través de Internet de material sujeto a derechos de autor, lo que tiene efectos perniciosos para las industrias de la creación, como el cine, la televisión, la edición, la música y el juego<sup>9</sup>.

En similar sentido, aunque un poco más preciso, el informe de la Oficina de Comercio de los Estados Unidos, sobre Piratería y Mercados notorios, concibe la piratería en línea como la práctica de “cargar, descargar, distribuir o transmitir por

Internet obras digitales protegidas por derechos de autor sin el permiso de los titulares de los derechos u otra autorización legal”<sup>10</sup>. De esta forma se han aceptado nuevas expresiones del fenómeno, como son la distribución de contenidos en plataformas sin licencia y demás supuestos de piratería facilitada por los servicios en línea, entre otros<sup>11</sup>.

Sin embargo, tanto en el supuesto de la piratería física como en línea, existen dos elementos esenciales que la determinan: el jurídico y el económico. El primero está dado por la infracción o violación de los derechos de autor, esencialmente, por la existencia de copias o reproducción de las obras; aunque dado los medios en que tiene lugar existe la infracción de otros derechos como el de distribución y puesta a disposición. El segundo, estrechamente relacionado con el primero, se delimita a partir de los tipos de bienes que son objetos de la piratería, así como el medio que se utiliza para ocasionarla, incluyendo el lugar en el que ocurre. Todo ello implica daños y perjuicios de naturaleza material, que podrían ser determinados a partir de los probables beneficios económicos que debían existir de haber sido las obras distribuidas y/o comercializadas por los canales legalmente establecidos.

Hasta este momento la relación que se ha establecido entre la inteligencia artificial y la piratería se ha sustentado, esencialmente, a partir de la utilización de esta tecnología para examinar textos e imágenes, en busca de pruebas de falsificaciones. Las plataformas de servicios de internet han utilizado la tecnología del aprendizaje automático “como forma de ampliar y adaptar rápidamente las medidas tradicionales de lucha contra la falsificación, como el filtrado de texto e imágenes”<sup>12</sup>. Sin embargo, todo parece indicar que esta relación pudiera cambiar. La Asociación de la Industria Discográfica de Estados Unidos (RIAA, por sus siglas en inglés), en su revisión de los mercados notorios de falsificación y piratería, presentado en octubre de 2022, identificó, sin denominarlos expresamente como piratas, a los denominados “extractores/mezcladores basados en IA”. Para dicha asociación, en este informe, estos agentes “extraen, o mejor dicho, copian, las voces, la música instrumental, o alguna parte de la música instrumental (una raíz musical) de una grabación de sonido, y/o generan, masterizan o remezclan una grabación para que sea muy similar o

6 OMC, disponible en [https://www.wto.org/spanish/thewto\\_s/glossary\\_s/piracy\\_s.htm](https://www.wto.org/spanish/thewto_s/glossary_s/piracy_s.htm), consultado el 16 de febrero de 2023.

7 STERLING, cit. pos. PANETHIERE (2005), pág. 2.

8 FREITAS (1992), pág. 3.

9 INTERPOL, disponible en <https://www.interpol.int/es/Delitos/Productos-ilegales/Compre-de-forma-segura/Pirateria-digital>, consultado el 16 de febrero de 2023.

10 OFFICE OF THE UNITED STATES TRADE REPRESENTATIVE, (2022B), 2022 Review of Notorious Markets for Counterfeiting and Piracy, Executive Office of the President, disponible en [https://ustr.gov/sites/default/files/2023-01/2022%20Notorious%20Markets%20List%20\(final\).pdf](https://ustr.gov/sites/default/files/2023-01/2022%20Notorious%20Markets%20List%20(final).pdf), consultado el 20 de febrero de 2023, pág. 4

11 La piratería de discos ópticos, la posesión de los dispositivos de piratería, los servicios ilícitos de televisión por protocolo de Internet (IPTV) que retransmiten ilegalmente señales de contenidos protegidos, el robo de señales de los operadores de cable, los “camcords” (grabaciones no autorizadas de películas recién estrenadas), la retransmisión no autorizada de programación deportiva en directo en línea; la clonación no autorizada de software de entretenimiento, piratería informática en servidores con contenidos pirateados en línea, y la distribución en línea de software y dispositivos que permiten eludir las medidas tecnológicas de protección. OFFICE OF THE UNITED STATES TRADE REPRESENTATIVE, (2022B), 2022 Review of Notorious Markets for Counterfeiting and Piracy, Executive Office of the President, disponible en [https://ustr.gov/sites/default/files/2023-01/2022%20Notorious%20Markets%20List%20\(final\).pdf](https://ustr.gov/sites/default/files/2023-01/2022%20Notorious%20Markets%20List%20(final).pdf), consultado el 20 de febrero de 2023, DALUGDUG (2022).

12 OFFICE OF THE UNITED STATES TRADE REPRESENTATIVE, (2022B), 2022 Review of Notorious Markets for Counterfeiting and Piracy, Executive Office of the President, disponible en [https://ustr.gov/sites/default/files/2023-01/2022%20Notorious%20Markets%20List%20\(final\).pdf](https://ustr.gov/sites/default/files/2023-01/2022%20Notorious%20Markets%20List%20(final).pdf), consultado el 20 de febrero de 2023, pág. 20.

casi tan buena como las pistas de referencia de artistas seleccionados y conocidos de grabaciones de sonido”<sup>13</sup>.

Pese a que finalmente la Oficina de Comercio de los Estados Unidos, no identificó a estos agentes como piratas, ni distribuidores de contenidos sin licencia<sup>14</sup>, la aparición en dicho informe no deja de ser llamativa. Expresión de un fenómeno que comienza a ser evidente, el uso sin autorización de obras para el entrenamiento de los sistemas de IA. A la declaración de la RIAA se le suman las demandas presentadas ante tribunales estadounidenses a inicios del año 2023 las cuales, más allá del resultado que obtengan, exponen el impacto que tiene el uso de obras como fuente de los datos para entrenar sistemas de IA generativos.

Las demandas revisten además, especial interés, porque aún cuando hacen alusión a un mismo hecho jurídico, como es el uso de las obras para entrenar sistemas de IA generativos, reflejan de manera distinta, el impacto de dicha conducta en los diversos titulares de los derechos de autor. Si bien en ambos casos las demandas son interpuestas por titulares de derechos y sus intereses se ven lacerados por el uso de las obras en el entrenamiento de estos sistemas, la posición que ostentan en el ámbito de la cadena de valor es diferente.

En la demanda colectiva los demandantes son los autores, los creadores de las imágenes, titulares originarios de los derechos de autor, mientras que, Getty Images es, esencialmente, una titular derivada cuyos derechos han sido adquiridos mediante licencias. Esta diferencia en las posiciones no solo permite sostener argumentos distintos en sus demandas, sino que evidencian de manera diferente el impacto de esta conducta y los daños económicos que producen. De ahí que, mientras los autores tomen como argumento la apropiación de los derechos de autor, así como la identidad y su prestigio artístico; Getty Images se basa, entre otras razones, en el valor que adquieren las imágenes como fuentes de datos de calidad, así como en las pérdidas de oportunidad en la distribución de contenidos digitales de calidad, inclusive, en las afectaciones que provoca en su modelo de negocios de licenciar “sus activos visuales y metadatos asociados en relación con el desarrollo de herramientas de inteligencia artificial y aprendizaje automático”<sup>15</sup>.

El impacto del uso de obras protegidas por los derechos de autor para entrenar sistemas de inteligencia artificial ya había sido abordado y debatido en algunos instrumentos internacionales. Por ejemplo, la Resolución del Parlamento Europeo, de 20 de octubre de 2020, sobre los derechos de propiedad intelectual para el desarrollo de las tecnologías relativas a la inteligencia artificial y la Recomendación sobre la ética de la inteligencia artificial, aprobada por aclamación en la 41 Conferencia General de la UNESCO, el 23 de noviembre de 2021. En la Resolución del Parlamento Europeo se hace referencia al uso de datos no personales por parte de la IA y, en particular, “la utilización lícita de obras y otras prestaciones protegidas por derechos de autor, así como de los datos —incluidos los contenidos preexistentes—, conjuntos de datos y metadatos conexos”<sup>16</sup>. En el ámbito de actuación número 7, referido a la Cultura, la Recomendación de la UNESCO hace un llamado a los Estados miembros a evaluar “cómo afectan las tecnologías de la IA a los derechos o los intereses de los titulares de derechos de propiedad intelectual cuyas obras se utilizan para investigar, **desarrollar, entrenar o implantar** aplicaciones de IA”<sup>17</sup>. La última versión del *Artificial Intelligence Act* de 2023, reconoce la obligación de los proveedores de *foundations models* utilizados en sistemas de IA generativos de “documentar y poner a disposición del público un resumen suficientemente detallado del uso de los datos de formación protegidos por la legislación sobre derechos de autor”, sin perjuicio de la legislación nacional o de la Unión sobre derechos de autor. Aunque no es el texto definitivo, no hay dudas que el reconocimiento de esta obligación, de conjunto con otras, es muestra del interés que adquiere la relación entre las IA generativas y la protección de las obras que se utilizan en el entrenamiento de los modelos. En particular, como reconoce el propio borrador, en la “generación de contenidos que infrinjan el Derecho de la Unión las normas sobre derechos de autor y el posible uso indebido”<sup>18</sup>.

Ante este complejo panorama surge entonces la necesidad de poder determinar si el uso de obras en el entrenamiento de los sistemas de IA generativos puede ser considerado una infracción de los derechos de propiedad intelectual de sus

13 RIAA (2022), Submission to Comment Request for the 2022 Review of Notorious Markets for Counterfeiting and Piracy Docket, No. USTR-2022-0010, disponible en <https://www.documentcloud.org/documents/23171871-ustr-2022-riaa>, consultado el 20 de enero de 2023; DALUGDUG (2022).

14 Entre los principales métodos se reconocen, la distribución de contenidos en streaming sin licencia, a través de aplicaciones, Stream ripping (“de-streaming”), la publicación, intercambio y descarga de archivos de contenidos a través de redes P2P (peer-to-peer), cyberlockers y servicios en la nube y las subastas en líneas. OFFICE OF THE UNITED STATES TRADE REPRESENTATIVE, (2022B), 2022 Review of Notorious Markets for Counterfeiting and Piracy, Executive Office of the President, disponible en [https://ustr.gov/sites/default/files/2023-01/2022%20Notorious%20Markets%20List%20\(final\).pdf](https://ustr.gov/sites/default/files/2023-01/2022%20Notorious%20Markets%20List%20(final).pdf), consultado el 20 de febrero de 2023, pág. p. 4

15 GETTY IMAGES (US), Inc. v. STABILITY AI, Inc. Demand for Jury Trial, United States District Court for The District of Delaware, Case 1:23-cv-00135-UNA, disponible en <https://www.courtlistener.com/docket/66788385/getty-images-us-inc-v-stability-ai-inc/>, consultado el 15 de febrero de 2023.

16 PARLAMENTO EUROPEO (2020), Resolución de 20 de octubre de 2020, sobre los derechos de propiedad intelectual para el desarrollo de las tecnologías relativas a la inteligencia artificial, (2020/2015(INI), disponible en [https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:JOC\\_2021\\_404\\_R\\_0007&from=EN](https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:JOC_2021_404_R_0007&from=EN), consultado el 15 de febrero de 2023, apartado 17.

17 UNESCO (2022), Recomendación sobre la ética de la inteligencia artificial, aprobada por aclamación en la 41 Conferencia General de la UNESCO, el 23 de noviembre de 2021, UNESCO, disponible en [https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa), consultado el 12 de enero de 2020, párrafo 99, (el resaltado es nuestro).

18 PARLAMENTO EUROPEO (2023), Proyecto de enmiendas a la Propuesta de Reglamento del Parlamento Europeo y del Consejo sobre normas armonizadas relativas a la inteligencia artificial (Ley sobre Inteligencia Artificial) y por el que se modifican determinados actos legislativos de la Unión, disponible en [https://www.europarl.europa.eu/meetdocs/2014\\_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA\\_IMCOLIBE\\_AI\\_ACT\\_EN.pdf](https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf), consultado el 9 de mayo de 2023, artículo 28.8.b).4c).

titulares, y si se debería denominar a esta infracción técnicamente, como piratería. Desde la perspectiva que sostenemos y argumentamos sí lo es. Todo ello si se toma como eje de perspectiva el hecho de que la piratería es un fenómeno que no permanece estático, sino que se desarrolla a partir de múltiples factores como son los bienes que se piratean, los medios y las formas que se utilizan para desarrollar esta actividad. Al analizar la relación que existe entre el uso de las obras, la tecnología de la IA y los modelos de negocios que se desarrollan alrededor de esta, es decir, cómo se comercializa, se concluye prima facie que sí es posible concebir al uso de las obras en el entrenamiento de los sistemas de IA generativos como un supuesto de piratería digital. Para poder explicar esta afirmación abordaremos los dos elementos en los que se sustenta, desde un punto de vista teórico, el fenómeno de la piratería, el legal y su impacto económico. Este último explicado a partir de la sustitución que el fenómeno supone.

## OBRAS, TECNOLOGÍAS Y DERECHOS DE AUTOR

La RIAA consideró, en el informe anteriormente referido, que la utilización de la música en el entrenamiento de los modelos de IA constituye un uso no autorizado e infringe los derechos de autor al realizar copias no autorizadas de las obras. A su juicio, los archivos que se difunden por estos servicios constituyen “copias no autorizadas o trabajos derivados no autorizados de la música de nuestros miembros”<sup>19</sup>. Por otra parte, en las demandas anteriormente mencionadas se identifican, al menos, tres grandes objetos de protección de los derechos de autor que son objeto de infracción, en mayor o menor medida. Las obras propiamente dichas, las bases de datos en las que están contenidas dichas obras y los sitios web.

En relación con las obras, en ambas demandas, por ejemplo, se señalan como conductas infractoras la reproducción, distribución, modificación y puesta a disposición sin autorización de las obras. Ello incluye la reproducción de la imagen en los ordenadores para el entrenamiento del modelo, su compresión en versiones o archivos más pequeños, con pérdida de detalles y el degradado para su entrenamiento, todo lo cual afecta la integridad de las obras. En relación con las bases de

datos se hace alusión a la extracción de su contenido, así como a la infracción de los sistemas de gestión de la información, los cuales permiten identificar las obras y crear metadatos, que supone una inversión cuantiosa, al propio tiempo que facilita su uso por los modelos de entrenamiento y mejora la calidad de las imágenes generadas por los sistemas de IA. Este ha sido, por ejemplo, uno de los argumentos más fuertemente utilizados por Getty Images en su demanda<sup>20</sup>.

Por último, y en relación con los sitios web, ha tenido lugar el proceso denominado “*Web Scraping*”, referido a la recolección, copia o extracción de datos de sitios web, utilizando herramientas automatizadas, como son los bots o rastreadores web. El denominado también como “raspados masivos” incluye no solo imágenes, sino también textos y códigos de software que se encuentren disponibles en el sitio web, así como cualquier otro bien jurídico que sea susceptible de protección por los derechos de propiedad intelectual. Esta actividad no solo está relacionada con una posible infracción de los derechos de autor desde la perspectiva de las leyes, sino también, de lo que normalmente prohíben los sitios web a partir de sus términos y condiciones<sup>21</sup>.

Sin embargo, debe tenerse en cuenta que al hablar del uso de las obras como fuente de datos para el entrenamiento de los sistemas de inteligencia artificial generativos la infracción de cada uno de los derechos de autor dependerá concretamente del modelo de generación utilizado, es decir de la tecnología. Por ende, no podemos afirmar que las infracciones sean siempre las mismas, o sigan la misma secuencia de pasos lógicos<sup>22</sup>, lo que si es posible precisar es que existe un uso no autorizado de grandes cantidades de obras como fuente de los datos de entrenamiento.

La IA generativa, según una IA generativa, el GhatGPT, no es más que “el conjunto de técnicas y algoritmos utilizados en la inteligencia artificial para crear contenido original, como imágenes, música, texto o videos, en lugar de simplemente realizar tareas basadas en reglas o patrones preexistentes”<sup>23</sup>. La idea principal de este concepto no es muy alejada de la que se sostiene en la última versión del borrador de la *Artificial Intelligence Act* de 2023, al referirse como tal a los “*foundations models*” destinados específicamente a generar, con distintos

19 RIAA (2022), Submission to Comment Request for the 2022 Review of Notorious Markets for Counterfeiting and Piracy Docket, No. USTR-2022-0010, disponible en <https://www.documentcloud.org/documents/23171871-ustr-2022-riaa>, consultado el 20 de enero de 2023; DALUGDUG (2022).

20 GETTY IMAGES (US), Inc. v. STABILITY AI, Inc. Demand for Jury Trial, United States District Court for The District of Delaware, Case 1:23-cv-00135-UNA, disponible en <https://www.courtlistener.com/docket/66788385/getty-images-us-inc-v-stability-ai-inc/>, consultado el 15 de febrero de 2023.

21 GETTY IMAGES (US), Inc. v. STABILITY AI, Inc. Demand for Jury Trial, United States District Court for The District of Delaware, Case 1:23-cv-00135-UNA, disponible en <https://www.courtlistener.com/docket/66788385/getty-images-us-inc-v-stability-ai-inc/>, consultado el 15 de febrero de 2023 y SARAH ANDERSEN, an individual; KELLY MCKERNAN, an individual; KARLA ORTIZ, an individual, Individual and Representative Plaintiffs, v. STABILITY AI LTD., a UK corporation; STABILITY AI, INC., a Delaware corporation; MIDJOURNEY, INC., a Delaware corporation; DEVIANTART INC, a Delaware corporation, Class Action Complaint, United States District Court Northern District Of California San Francisco Division, Case 3:23-cv-00201, disponible en <https://stablediffusionlitigation.com/pdf/00201/1-1-stable-diffusion-complaint.pdf>, consultado el 20 de enero de 2023.

22 Por ejemplo, en el modelo de difusión existe una compresión con pérdidas de la imagen, y al propio tiempo ocurre su reconstrucción, sin que exista una coincidencia entre las utilizadas para su entrenamiento y el resultado, excepto algunos casos. Empero, como lo han resumido algunos autores, “los modelos generativos pueden verse como una copia comprimida y organizada de un conjunto de datos”. JAHANIAN, PUIG, TIAN, ISOLA (2022), pág. 3.

23 <https://chat.openai.com/chat>. El uso de este sistema de inteligencia artificial solo ha sido para demostrar como el propio sistema se describe.

niveles de autonomía, contenidos como texto complejo, imágenes, audio o vídeo (“IA generativa”)<sup>24</sup>.

Utilizada para generar contenidos creativos a partir de redes neuronales y algoritmos de aprendizaje, lo cierto es que esta tecnología se caracteriza por la variedad de tipos de redes y modelos que utiliza. Estos evolucionan constantemente. En la actualidad, por ejemplo, los sistemas de IA generativos de empresas como Dall-e de OpenAI, Stability y Midjourney, entre otros, se han desarrollado a partir de modelos de difusión<sup>25</sup> y de comprensión del lenguaje, que no solo son capaces de crear imágenes de alta calidad, sino también de crearlas a partir de instrucciones concretas que reciban estos sistemas.

Tanto en los modelos de difusión como en cualquier otro, existe una estrecha relación entre las obras que son utilizadas para su entrenamiento (*inputs*) y el resultado creativo que se obtiene de su aplicación (*outputs*). Por ende, las obra y su calidad, como fuente de datos de entrenamientos para el resultado que finalmente va a arrojar la IA generativa es definitoria. Aun cuando se hace referencia al término de obras, lo cierto es que el uso de datos de entrenamiento no solo implica el uso de estas, cualquiera que sea su tipo, también podrán ser utilizadas como fuentes, otras prestaciones artísticas protegidas por los derechos de propiedad intelectual, como pueden ser las interpretaciones y ejecuciones de los artistas, incluyendo del ámbito audiovisual como musical. En estos últimos casos, están presentes otros bienes jurídicos merecedores de tutela jurídica como son la voz y la imagen de los artistas<sup>26</sup>.

En cualquier caso, lo cierto es que los sistemas de IA generativos no pueden superar las “limitaciones de las imágenes de entrenamiento”<sup>27</sup>, en otras palabras, las limitaciones de las obras y prestaciones artísticas utilizadas. Hasta este momento, sea cual sea el modelo generativo desarrollado, la tecnología de IA tiene limitaciones intrínsecas, en particular cómo opera el proceso de “creatividad”. Todo ello se ha reconocido por expertos en la materia. JAHANIAN, PUIG, TIAN e ISOLA, comentan, “las representaciones resultantes rivalizan con las obtenidas directamente a partir de datos reales, o incluso las superan” (...), “pero para obtener buenos resultados es ne-

cesario de la estrategia de muestreo aplicada y del método de entrenamiento”<sup>28</sup>. Uno de los fundadores de las empresas que desarrollaron algunos de estos sistemas de IA describe de forma muy clara como operan. “Cuando se le solicita a la IA que realice un dibujo, en realidad esta no recuerda ni sabe nada de lo que ha hecho. No tiene voluntad, no tiene objetivos, no tiene intención, no tiene capacidad de contar historias” (...) “Es como un motor. Un motor no tiene adónde ir, pero la gente tiene adónde ir”<sup>29</sup>.

Es precisamente este carácter técnico y limitado lo que refuerza el valor que adquieren los datos y las obras que los contienen en el desarrollo de esta tecnología y los resultados que arrojan. De ahí que, aun cuando se hable de “inspiración”, “aprendizaje”, “capacidad de entender”, los sistemas de IA generativos responden a la lógica de los métodos estadísticos y matemáticos, es decir, no se inspiran, razonan u aprenden como los seres humanos. Dependen tanto de los modelos algorítmicos que se utilizan, como de los datos de entrenamiento, léase en este caso, las obras. Pero incluso, aun en estas circunstancias es en la fuente de los datos y su disponibilidad donde recae el verdadero valor de estos sistemas generativos, y las obras son precisamente eso, fuente de datos. El futuro, tal como lo prevén JAHANIAN, PUIG, TIAN e ISOLA será el de la proliferación de “zoológicos de modelos”, “mientras los conjuntos de datos se vuelven cada vez más difíciles de manejar, inexistentes o privados”<sup>30</sup>.

Conforme el desarrollo y avance de la tecnología, identificar cuál o cuáles son las infracciones que se cometen a partir del uso de las obras para entrenar sistemas de IA generativos no es relevante, ni desde un punto de vista jurídico ni tecnológico. Más allá de los artificios tecnológicos que se pudieran llevar a cabo para poder lograr que las obras sean utilizadas como datos de entrenamiento, y que pudieran significar una infracción de los derechos de autor, lo verdaderamente trascendente es que existe un uso no autorizado de estas obras, para las cuáles no fueron concebidas y que, este uso, no cuenta con autorización. En cualquier caso como fuente de datos de entrenamiento estas tendrán que ser modificadas, reproduci-

24 PARLAMENTO EUROPEO (2023), Proyecto de enmiendas a la Propuesta de Reglamento del Parlamento Europeo y del Consejo sobre normas armonizadas relativas a la inteligencia artificial (Ley sobre Inteligencia Artificial) y por el que se modifican determinados actos legislativos de la Unión, disponible en [https://www.europarl.europa.eu/meetdocs/2014\\_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA\\_IMCOLIBE\\_AI\\_ACT\\_EN.pdf](https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf), consultado el 9 de mayo de 2023, artículo 28.b).

25 Son diversos los métodos que se utilizan para generar imágenes, además de los modelos de difusión existen, por ejemplo, los codificadores automáticos y las redes generativas adversariales (GAN). Entre estos modelos existen diferencias en términos de diversidad y calidad, e inclusive en la forma como opera la tecnología. Por ejemplo, las GAN utilizan dos redes neuronales que compiten entre sí (adversarial), mientras que en los modelos de difusión solo interviene una, en un proceso que se denomina ruido, en el cual se reconstruyen las copias de las imágenes utilizadas en el entrenamiento. Vid. CARR (2022).

26 Por ejemplo, en el sistema Lingyin de TME se utiliza una tecnología de síntesis de voz que permite replicar las voces de los cantantes para producir canciones originales en cualquier estilo e idioma. Ello ha incluso posibilitado el uso de voces sintéticas en memorias de artistas legendarias incluso fallecidas. La empresa HYBE por ejemplo, recreó la voz de una superestrella surcoreana del folk, ya fallecida (Kim Kwang-seok). Esto evidentemente supone nuevos retos. STASSEN (2022).

27 SARAH ANDERSEN, an individual; KELLY MCKERNAN, an individual; KARLA ORTIZ, an individual, Individual and Representative Plaintiffs, v. STABILITY AI LTD., a UK corporation; STABILITY AI, INC., a Delaware corporation; MIDJOURNEY, INC., a Delaware corporation; DEVIANTART INC, a Delaware corporation, Class Action Complaint, United States District Court Northern District Of California San Francisco Division, Case 3:23-cv-00201, disponible en <https://stablediffusionlitigation.com/pdf/00201/1-1-stable-diffusion-complaint.pdf>, consultado el 20 de enero de 2023, apartado 98.

28 Jahanian, Puig, Tian, Isola (2022), pág. 9.

29 Vincent (2022).

30 JAHANIAN, PUIG, TIAN, ISOLA (2022), pág. 1.

das, puestas a disposición, etc. Todo lo cual impacta directa y negativamente en los beneficios económicos que pudieran obtener los autores y demás titulares por la explotación de sus derechos, es decir, provoca un marcado perjuicio económico.

## 1. La apreciación de limitaciones o excepciones a los derechos de autor

Si se concibe que el uso sin autorización de las obras para el entrenamiento de los sistemas de IA generativos es una infracción de los derechos de autor, deberíamos también preguntarnos si dicho uso pudiera ser considerado una excepción o limitación en el ejercicio de los derechos patrimoniales de autor. La Resolución del Parlamento Europeo, anteriormente citada, refiere la necesidad de que el uso lícito de obras y prestaciones protegidas por los derechos de autor, por parte de los sistemas de IA sea evaluado a la luz “de las normas existentes sobre las limitaciones y excepciones a la protección mediante derechos de autor, como la prospección de texto y datos”<sup>31</sup>.

En correspondencia con lo señalado por el Parlamento Europeo, algunos autores hacen referencia a que podría ser entendido o considerado un supuesto de *fair use*. Guadamuz, por ejemplo, refiere que dada la naturaleza técnica se podría aplicar la excepción de copia individual o la excepción de la minería de textos y datos. En el primer supuesto el autor considera que la copia individual de cada obra utilizada en el entrenamiento pudiera no tener “importancia económica independiente”; mientras que, en el segundo caso, se trataría de la regulación de cada ordenamiento jurídico y los fines de su reconocimiento, que varían según cada país<sup>32</sup>.

Sin embargo, hay que tener en cuenta que no coincide técnicamente la minería de textos y datos, con el uso de las obras y demás prestaciones artísticas en el entrenamiento de los sistemas de IA generativos. Sin entrar a dilucidar entre los beneficios y aplicaciones de uno u otro caso debe tenerse en cuenta que: la minería de textos y datos, se define en la Directiva (UE) 2019/790 del Parlamento Europeo y del Consejo, sobre los derechos de autor y derechos afines en el mercado único digital, como aquella “técnica analítica automatizada

destinada a analizar textos y datos en formato digital a fin de generar información que incluye, sin carácter exhaustivo, pautas, tendencias o correlaciones”<sup>33</sup>. Constituye una técnica de investigación computacional que se puede utilizar sobre obras protegidas por los derechos de autor, pero también en otras fuentes de datos. Como bien señala Carroll los resultados de su utilización “informan de correlaciones, patrones u otras relaciones encontradas en la información que se ha extraído”, por ende, tiene un impacto en la investigación e innovación<sup>34</sup>.

Para que tenga lugar el aprendizaje automático podría ser necesario el uso de minería de textos y datos, pero, ello no tiene que ser necesariamente así, en particular cuando se hace referencia a sistemas de IA generativos desarrollados a partir de modelos de difusión. Tampoco no todos los usos de minería de textos y datos tienen como objetivo enseñar a los sistemas de IA a que aprendan y, en el caso de que su fin fuera este, los sistemas no tienen porque ser generativos. La posibilidad de que el entrenamiento de datos pueda ser incluido dentro de la excepción de la minería de textos y datos de obras se encuentra en dependencia de la regulación que al respecto realice cada país. Aunque ha de tenerse en cuenta que los procedimientos tecnológicamente no coinciden y dependen en gran medida del tipo de tecnología. Por otro lado, no puede olvidarse que la interpretación de los usos y limitaciones es restrictiva, y que el uso de obras y otras prestaciones artísticas no puede provocar un perjuicio a los autores y demás titulares.

Más allá de la aplicación del sistema de excepciones y limitaciones continental o del *fair use*, según el sistema de cada país, lo cierto es que, el uso de obras para entrenar sistemas de IA generativos incumple uno de los requisitos sustanciales de ambos regímenes, no causar un perjuicio injustificado a los intereses legítimos del autor<sup>35</sup> o, su utilización conforme el impacto sobre el mercado potencial o el valor de la obra protegida<sup>36</sup>. De hecho, los mismos argumentos que enervan, en principio, cualquier posibilidad de aplicar un régimen de limitaciones y excepciones o, el del *fair use*, son los mismos que nos hacen argumentar que existe piratería.

31 PARLAMENTO EUROPEO (2020), Resolución de 20 de octubre de 2020, sobre los derechos de propiedad intelectual para el desarrollo de las tecnologías relativas a la inteligencia artificial, (2020/2015(INI), disponible en [https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:JOC\\_2021\\_404\\_R\\_0007&from=EN](https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:JOC_2021_404_R_0007&from=EN), consultado el 15 de febrero de 2023, apartado 17.

32 GUADAMUZ (2023), págs. 14 y ss.

33 PARLAMENTO EUROPEO Y DEL CONSEJO (2019), Directiva (UE) 2019/790 de 17 de abril de 2019, sobre los derechos de autor y derechos afines en el mercado único digital y por la que se modifican las Directivas 96/9/CE y 2001/29/CE, disponible en <https://eur-lex.europa.eu/legal-content/es/TXT/?uri=CELEX:32019L0790>, consultado el 20 de febrero de 2023, art. 2.2.

34 Carroll (2019), pág. 895.

35 El artículo 9.2 del Convenio de Berna, brinda la posibilidad de que los Estados de la Unión permitan la reproducción de las obras en casos especiales, siempre y cuando esta reproducción no atente a la explotación normal de la obra ni cause un perjuicio injustificado a los intereses legítimos del autor. UNIÓN DE BERNA, (1886), Convenio de Berna para la protección de las Obras Literarias y Artísticas, enmendado el 28 de septiembre de 1979, disponible en <https://www.wipo.int/wipolex/es/text/283700>, consultado el 15 de febrero de 2023.

36 En el sistema de *fair use* también se establecen requisitos que deben ser tenidos en cuenta para su aplicación, dentro de los que se encuentran (1) el propósito y el carácter del uso, incluyendo si dicho uso es de naturaleza comercial o educativa sin ánimo de lucro; (2) la naturaleza de la obra protegida por derechos de autor; (3) la cantidad y sustancialidad de la parte utilizada en relación con la obra protegida en su conjunto; y (4) el efecto del uso sobre el mercado potencial o el valor de la obra protegida. UNITED STATES COPYRIGHT OFFICE, (2022), Copyright Law of the United States and Related Laws Contained in Title 17 of the United States Code, United States Copyright Office, October, disponible en <https://www.copyright.gov/title17/title17.pdf>, consultado el 15 de febrero de 2023, §107 · Limitations on exclusive rights: Fair use.

## EL IMPACTO ECONÓMICO, LA OTRA CARA DE LA PIRATERÍA

Como ya se ha referido, al abordar el tema de la piratería no solo se debe tomar en cuenta las infracciones de los derechos que existen por el uso de las obras, sino también, el impacto que dicha conducta pudiera ocasionar, tanto para los autores y demás titulares de los derechos de autor como los beneficios que supone para los desarrolladores de este modelo de negocios. Ambas vertientes serán analizadas a continuación.

### 1. Los beneficios para los desarrolladores de los sistemas de IA generativos

Pudiera pensarse, en principio, que este tipo de uso sin autorización se realiza de buena fe. De hecho, David Holz, uno de los creadores de Midjourney ha afirmado que los datos de entrenamiento se han extraído del mismo lugar de donde se extraen el resto de los datos, Internet. En este sentido argumenta que, siendo una ciencia experimental que aún se encuentra en fase temprana de desarrollo “se ha accedido a lo que se puede, sin saber si los datos a los que se accede son realmente importantes”<sup>37</sup>. Y no hay dudas de que ello pudiera ser así, pues no es menos cierto que el desarrollo de estos modelos no solo es nuevo, sino también, costoso. El problema se halla en determinar en qué momento deja de ser considerado un modelo de desarrollo experimental para convertirse en un modelo exitoso que genera ganancias para sus creadores y gana aceptación por parte de los consumidores.

En la demanda colectiva de Sarah Andersen, Kelly McKernan y Karla Ortiz y otros, se devela otra arista de este fenómeno que debe ser tenida en cuenta. No es solo una cuestión de creencia, concebir que todo lo que se encuentra en Internet está abierto, lo que es lógico que no es así; sino más bien, los costos que supone poder recabar la autorización de los titulares de derechos para poder tener a su disposición las obras protegidas por los derechos de autor que son necesarias para el entrenamiento del modelo. El dilema no se encuentra, como bien se afirma en la demanda, en que no sea posible rastrear los titulares de estas imágenes, sino en lo costoso que ello sería<sup>38</sup>, lo cual es diferente.

Ello devela la interrogante de si los derechos de autor podrían constituirse en una barrera a la innovación y un límite al desarrollo de estos sistemas de IA. En una fase experimental es lógico que los costos de transacción que ello supone no podrían ser asumidos por este tipo de “emprendimientos”, los

Como ha reconocido la Suprema Corte de los Estados Unidos, la determinación del perjuicio y del impacto potencial y/o real del uso de las obras sin autorización, para poder aplicar la doctrina del fair use, precisa del análisis de cuestiones de hecho y no solo estrictamente jurídicas. La determinación de qué cantidad de obras protegidas por los derechos de autor es utilizada y cuál es el daño que se produce en los mercados reales y potenciales de estas obras, es sin dudas una cuestión fáctica, que ciertamente existe, como se analizará oportunamente, aunque en ocasiones es difícil de determinar o probar.

En la apreciación del régimen de limitaciones o excepciones la finalidad comercial quizás no sea de especial relevancia, en particular, al momento de evaluar el impacto de esta conducta en las obras y demás prestaciones pirateadas. Ello obedece a que el fin comercial no siempre queda delimitado en el tipo de modelo de negocios que se desarrolla a partir de estos sistemas. Por ejemplo, algunos ofrecen modelos gratuitos y accesos premium bajo pago<sup>37</sup>. Mientras que, como bien ha reconocido la propia Suprema Corte de los Estados Unidos, al interpretar los usos permitidos bajo el fair use, estos no tienen que ser necesariamente no comerciales, siempre y cuando el propósito del uso sea verdaderamente transformativo, añada algo nuevo e importante y contribuya a fomentar la creatividad<sup>38</sup>.

En cualquier caso, como suele ocurrir en materia de limitaciones y excepciones a los derechos de autor, serán las regulaciones nacionales quienes finalmente decidan al respecto sobre la consideración o no de este uso como limitación o excepción. Sin embargo, para su adopción siempre será necesario responder a preguntas como, ¿Cuáles son los beneficios que se obtendrían con ello y, en particular, con este uso no autorizado? ¿Estos beneficios implican un interés para la producción creativa de nuevas expresiones protegidas por los derechos de autor? ¿Las pérdidas económicas para los titulares de los derechos que dicho uso supone, real o potencialmente, son significativamente menores que la utilidad que se obtiene con su permisibilidad? ¿Quiénes serían los beneficiarios de este uso no autorizado? En el siguiente apartado explicaremos, aun sin la profundización necesaria, algunas de nuestras consideraciones sobre este tema.

37 Por ejemplo, Midjourney, permite realizar un registro de cuenta gratuita, obteniendo el usuario 25 créditos. Posteriormente puede pagar de 10 a 30 dólares al mes, lo que le posibilita tener mayor o menor acceso a imágenes públicas o privadas.

38 GOOGLE LLC v. ORACLE (2021), America Supreme Court Of The United States, Inc. No. 18-956, disponible en [https://www.supremecourt.gov/opinions/20pdf/18-956\\_d18f.pdf](https://www.supremecourt.gov/opinions/20pdf/18-956_d18f.pdf), consultado el 20 de julio de 2022.

39 Vincent (2022).

40 SARAH ANDERSEN, an individual; KELLY MCKERNAN, an individual; KARLA ORTIZ, an individual, Individual and Representative Plaintiffs, v. STABILITY AI LTD., a UK corporation; STABILITY AI, INC., a Delaware corporation; MIDJOURNEY, INC., a Delaware corporation; DEVIANTART INC, a Delaware corporation, Class Action Complaint, United States District Court Northern District Of California San Francisco Division, Case 3:23-cv-00201, disponible en <https://stablediffusionlitigation.com/pdf/00201/1-1-stable-diffusion-complaint.pdf>, consultado el 20 de enero de 2023.

que además realizan múltiples pruebas en busca de un constante perfeccionamiento. Un aspecto que no suele abordarse con frecuencia, pero que es importante tomar en consideración, es el de los costos de estos desarrollos y, en particular, aquellos que están asociados a los derechos de autor. Se calcula que, como promedio, entrenar estos modelos genera un gasto de alrededor de 50 mil dólares cada vez que se hacen, y se necesitan múltiples intentos. Asimismo, se precisan de equipos y servidores potentes para llevar cabo estos procedimientos en cortos períodos de tiempo<sup>41</sup>. Ante estas circunstancias es lógico que al buscar ahorrar, dentro del desarrollo de la tecnología, los más afectados sean precisamente los más débiles dentro de la cadena de valor de la misma, y estos, no son otros, que los autores, quienes aportan “la materia prima” fundamental de la tecnología y a partir de la cual se desarrolla el modelo de negocios.

Dado que el ahorro de los desarrolladores implica una pérdida para los autores y demás titulares, no puede desconocerse el salto que tiene lugar en la cadena de valor, entre los modelos de pruebas y los modelos de negocios que se desarrollan a partir del entrenamiento de obras protegidas por derechos de autor. Es precisamente en este punto de quiebre que se hace necesario recabar las correspondientes licencias y autorizaciones de uso de las obras y prestaciones protegidas. Dado que no es igual la tecnología en fase de prueba que los modelos de negocios que se consolidan económicamente y son rentables, podría incluso pensarse en posibles licenciamientos no voluntarios. Futuros estudios deberán analizar cómo implementar dichas licencias y autorizaciones, cuál es el monto que se deba pagar por las mismas y quienes lo deban hacer.

La contradicción que se genera entre autores, titulares y desarrolladores de sistemas de IA generativos ha sido presentada, aunque de manera indirecta por el propio Holz, de quien se ha hecho referencia anteriormente. Para el fundador de Midjourney, el tema del desarrollo de este tipo de sistemas tiene un marcado carácter económico, dado los costos que ello supone. “Será caro, pero será posible”, llegó a afirmar; sin embargo, en su misma opinión, pero refiriéndose a los autores, es una cuestión de “inspiración” y “reconocimiento”, en el que el artista quiere que la gente “se inspire en las cosas que hace”<sup>42</sup>. De esto último, no hay dudas, pero los autores y artistas también necesitan obtener frutos de su trabajo, de su

creatividad. Perspectivas como estas, solo contribuyen a incrementar la brecha de valor de los autores, encubriendo la realidad del problema, que no es otro que la distribución justa y equitativa de los beneficios económicos obtenidos entre todos los agentes que intervienen en la cadena de valor del negocio que se desarrolla a partir de las IA generativas.

Al igual que ha sucedido con otros intermediarios y plataformas digitales, la primera reacción cuando surgen es intentar desconocer el derecho de los creadores a obtener réditos de su creación, argumentando su papel “innovador” en el ecosistema digital, sin embargo, sin menoscabo del papel que desempeñan, no se puede ignorar que, hasta este momento, no dejan de ser intermediarios que usan la creatividad de los autores para brindar servicios económicamente rentables. Aunque, a diferencia de las plataformas y demás intermediarios que hoy existen en el ámbito digital, el impacto de las IA generativas es aun incierto. Este último puede ser mayor o menor no solo desde el ámbito de la piratería, es decir, del uso de las obras sin autorización por parte de estos servicios para entrenar sistemas, sino también, desde el propio ámbito de la creación y la sustitución de la creatividad humana. Las dimensiones de estas pérdidas se exponen a continuación.

## 2. Las pérdidas económicas de los creadores y demás titulares por el uso de sus obras

A diferencia de lo que suele pensarse la piratería no está necesariamente disociada de la fiabilidad de los contenidos y de su calidad, ello va de la mano con el aumento de la disponibilidad de contenidos y su atractivo<sup>43</sup>. Ambos elementos influyen en la decisión del consumidor para poder adquirir productos piratas. Los productos que generan los sistemas de IA creativos cumplen precisamente con estos criterios, la diferencia fundamental se encuentra en que el producto que arroja el sistema de IA creativo no es igual a la obra que ha sido pirateada, excepto en determinados supuestos particulares<sup>44</sup>. Por ello, en principio no se puede hablar de que, entre la obra cuyos derechos se infringen y la “obra” que se comercializa existe una identidad para denominarla copia, es decir, no existe la figura del sustituto tal como hasta ahora se reconoce en el ámbito de la piratería.

41 Vincent (2022).

42 Vincent (2022).

43 Danaher; Smith; Telang (2020), pág. 34.

44 Un reciente estudio realizado por investigadores de diversas instituciones confirmó la posibilidad de que algunas imágenes generadas por modelos de fusión coincidan con las imágenes originales individuales que se utilizaron en las bases de datos de entrenamiento del modelo. Esto es lo que se denomina memorización o regeneración de imágenes. Fueron al menos encontradas 100 réplicas de imágenes en el conjunto. Carlini; Hayes; Nasr; Jagielski; Sehwag; Tramèr; Balle; Ippolito; Wallace (2023); Heikkilä (2023), pág. 4. La obtención de copias exactas de imágenes y demás datos de entrenamiento por parte de modelos de difusión, que las memorizan y regeneran, guarda una mayor similitud con el plagio dado que se trata del output del sistema, generándose una creencia, de buena fe, de que dicho resultado es un “acto creativo” de este, y no la copia del original que ha sido utilizado como inputs para su entrenamiento. Si bien, desde una perspectiva probatoria en este supuesto es más fácil demostrar la infracción que implica el uso de la obra en el entrenamiento del sistema, se deberán tener en cuenta otras posibles infracciones que tienen lugar a partir de los usos que se realicen del resultado generado, incluyendo la creencia de una posible autoría que no ha tenido lugar. Este caso no es el objeto de este artículo.

El concepto de los sustitutos de las obras originales está estrechamente relacionado con el de la piratería, en particular por ser este la fuente de la pérdida de ingresos y de los daños relacionados con este fenómeno. Se entiende como tal aquellas copias de las obras que son comercializadas y/o distribuidas sin autorización de los titulares de los derechos de autor. Sin embargo, el concepto de sustitución de las obras aplicado al ámbito de la piratería de la IA, debe ser interpretado teniendo en cuenta las propias particularidades de la tecnología y su uso en el mercado. Para ello, deben considerarse dos elementos o factores, el uso que la tecnología hace de las obras para generar un resultado “creativo” y, el resultado final creativo en sí mismo, el sustituto de la obra utilizada para entrenar los sistemas generativos.

Un aspecto que necesita ser precisado es que el objeto de nuestro análisis no es el uso de las obras para entrenar cualquier sistema de IA, sino solo aquellos sistemas que se denominan generativos, y esto es así porque no todos los supuestos de uso de las obras para entrenar sistemas de IA deben ser entendidos como piratería. Para ello, debe de reconocerse que la existencia de este tipo de sistemas supone una “sustitución de la actividad creativa”, más allá de que se reconozcan derechos o no sobre dicha actividad y de la mayor o menor intervención de los seres humanos en el resultado. De facto, los outputs que arrojan dichos modelos impactan en la forma en la que tradicionalmente se ha concebido la creatividad. Por ello, cuando se analiza el fenómeno de la piratería de la IA no se puede identificar al sustituto en sí mismo en las obras que se utilizan para el entrenamiento, sino en el resultado de la generación de la IA que se desarrolla a partir de dicho uso.

Acá retomamos algo a lo que se había hecho referencia, la relación que existe entre el dato que sirve para el entrenamiento, las obras que son su fuente y el “resultado creativo”. La “sustitución de la creatividad” no puede tener lugar si no se utilizan las obras que sirven para entrenar dichos modelos, como se ha citado anteriormente, “las representaciones resultantes rivalizan con las obtenidas directamente a partir de datos reales, o incluso las superan”<sup>45</sup>. Empero, para que exista dicha rivalidad y/o superación es necesario contar con datos de calidad. El sustituto de la obra original que aparece en el mercado no puede existir sin la obra que se ha utilizado sin autorización, sin la fuente de la calidad de los datos de entrenamiento. En otras palabras, para generar productos “creativos de calidad” se necesitan “obras de entrenamiento de calidad”, lo cual, además, mejora ostensiblemente la explotación económica de los resultados generados.

Para que las “resultados creativos” de la IA puedan rivalizar, y sustituir las creaciones reales de personas físicas, nece-

sitan de datos de calidad, provenientes de fuentes fiables que permitan cumplir iguales funcionalidades; es decir, compiten con las mismas obras que han sido utilizadas para su entrenamiento y le dieron “vida”. Esto no quiere decir que el sustituto de la obra resultado de la IA deba ser considerado pirata, lo que debe considerarse como pirata es el uso de obras reales creadas por personas físicas que han sido utilizadas para producir el resultado sustituto.

Si se analiza la sustitución de las obras que tiene lugar a partir de su uso no autorizado por los sistemas de IA generativos, podremos percatarnos que el impacto de esta sustitución no se halla solo en relación con la obra propiamente dicha, sino también, en relación con los modelos de negocios que se desarrollan a partir de estas, e incluso, en su propia configuración como fuente de datos de entrenamiento. Por ello, si se examina el proceso tecnológico que ocurre en este tipo de sistemas de forma holística, incluyendo no solo su concepción, sino también, su puesta a disposición y los modelos de negocios que alrededor del mismo se sustentan, podremos percatarnos que el impacto del uso de obras no autorizadas para el entrenamiento de sistemas de IA es similar al que acontece en la piratería habitualmente, no solo desde un punto de vista jurídico (en cuanto a la violación de los derechos de autor), sino también, económico.

Para analizar el impacto económico debe considerarse que las pérdidas potenciales o supuestas para los titulares de derechos entran en conflicto con el propósito de los derechos de autor, “proporcionar a los autores derechos exclusivos que estimulen la expresión creativa”<sup>46</sup>. Para su determinación no solo se toma en cuenta la cantidad de dinero que el titular de los derechos podría perder, es decir, su cuantía, sino también, el origen de la pérdida, que en este caso se encuentra en el uso que se hace de las obras sin autorización para realizar el entrenamiento de los sistemas de IA. Dicho uso genera un resultado creativo, el sustituto, que compite con las obras realizadas por personas físicas comercializadas en un mismo mercado pero, además, en condiciones de desigualdad.

La desigualdad se representa desde un doble carácter, desde el punto de vista de la tecnología como desde la comercialización del producto. Desde la tecnología, la capacidad de la IA generativa de poner a disposición de los consumidores una gran cantidad de obras en breve *lapsus* de tiempo, pudiera poner a las personas físicas creadoras a competir con ellas para satisfacer la demanda de obras, en un mercado único, y con un costo de creación prácticamente cero, en comparación con el que implica para los creadores físicos; exceptuando, por supuesto, el costo que en sí mismo tiene el desarrollo de la tecnología, que es diferente del de la creación de la obra. Por

45 Jahanian, Puig, Tian, Isola (2022), pág.9.

46 GOOGLE LLC v. ORACLE (2021), America Supreme Court Of The United States, Inc. No. 18-956, disponible en [https://www.supremecourt.gov/opinions/20pdf/18-956\\_d18f.pdf](https://www.supremecourt.gov/opinions/20pdf/18-956_d18f.pdf), consultado el 20 de julio de 2022.

ende, el pago de los derechos de autor a los creadores físico no puede ser interpretado como una monopolización del mercado, que evita la competencia e impide la promoción de la expresión creativa, sino más bien, como un instrumento que permite una justa distribución del valor entre los diferentes actores que se desarrollan alrededor de la cadena de valor de la IA generativa.

Desde el punto de vista del mercado, la desigualdad se agrava no solo por el uso de la tecnología en un mercado único, sino también, por el hecho de que en este proceso de “aprendizaje”, las obras productos de los sistemas de IA generativos copian también el estilo de muchas de las obras originales que son utilizadas en el proceso de entrenamiento. Esto último, no solo refuerza la capacidad de sustitución de las obras originales por parte de las obras generadas por IA, sino que, particularmente, suponen también un aprovechamiento del valor comercial que pueden tener las obras que son protegidas por los derechos de autor, al crear confusión entre los consumidores.

Aun cuando el estilo, en sí mismo considerado no es objeto de protección por parte de los derechos de autor, sí pudiera entrar en consideración desde el ámbito de la competencia desleal y la confusión que pudiera provocar al consumidor. En la demanda colectiva, anteriormente referida, se precisa que los compradores que buscan una nueva imagen, “al estilo” de un artista determinado, ya no tendrán que pagar por encargar o licenciar una imagen original a este u otro artista, sino que, el sistema utilizando su nombre, así como las imágenes que fueron usadas en el entrenamiento, generaría nuevas obras con similar estilo, pero sin compensar el creador original, ni por el uso del nombre así como tampoco de sus imágenes originales. Como señala, Greg RUTKOWSKI cuyo peculiar estilo artístico es uno de los más utilizados para estos sistemas, en particular, Stable Diffusion, “ (...) seguramente no podré encontrar mi trabajo allí porque Internet se llenará de obras de arte creadas por la IA. Eso es preocupante”<sup>47</sup>. Desarrollar obras de arte que se ajusten a estilos particulares de un artista, es algo que, como también se afirma en dicha demanda, solo

es posible de realizar por un número muy reducido de artistas<sup>48</sup> (léase personas físicas).

Empero, como también se ha argumentado en las respectivas demandas, el uso de estas obras en el entrenamiento de estos modelos, no solo provoca una sustitución en cuanto a las obras, sino también, en relación con los modelos de negocios que se desarrollan alrededor de estas. Por ende, no puede decirse que el perjuicio económico de este acto recaerá solo en cabeza de los creadores, sino también, en el resto de los titulares derivados de los derechos patrimoniales de autor, como es el caso de Getty Images<sup>49</sup>; distribuidor de contenidos digitales y titular de derechos de autor de millones de activos visuales. Para estos titulares de derechos, las empresas desarrolladas a partir de modelos de IA generativos constituyen una competencia directa como distribuidor de contenidos, al propio tiempo que utilizan, sin su autorización, sus bases de datos para generar los contenidos que finalmente serán su competencia.

A la situación anteriormente descrita se le suma el hecho de que para generar esta competencia, ya sea en “la creatividad” como en los modelos de negocios, los sistemas de IA generativos utilizan para su entrenamiento otros bienes jurídicos como son los nombres de los creadores, marcas, entre otros bienes jurídicos tutelados<sup>50</sup>; sin los cuáles no se puede desarrollar el procesamiento del lenguaje natural que permite al sistema de IA asociar un determinado estilo creativo con un autor o creador en particular. Esta asociación, que posibilita el desarrollo de la tecnología y su uso de manera en la que actualmente la conocemos, ha sido interpretada en la demanda colectiva como una apropiación de las identidades artísticas distintivas que se generan a partir de la relación entre el nombre, el estilo artístico y la creatividad de los autores<sup>51</sup>, lo cual es cierto y no necesariamente fortuito, puesto que está determinado, en gran medida, desde el mismo proceso de selección y uso de las obras en el entrenamiento del modelo y la generación de valor a partir de productos creativos más acabados y artísticamente más aceptados.

Por último, y no menos importante en la determinación del impacto de este tipo de piratería debe reconocerse la susti-

47 Se considera que el nombre de este artista aparece 90.000 veces en Léxica, una web que rastrea imágenes generadas por IA. Heikkilä (2022).

48 SARAH ANDERSEN, an individual; KELLY MCKERNAN, an individual; KARLA ORTIZ, an individual, Individual and Representative Plaintiffs, v. STABILITY AI LTD., a UK corporation; STABILITY AI, INC., a Delaware corporation; MIDJOURNEY, INC., a Delaware corporation; DEVIANTART INC, a Delaware corporation, Class Action Complaint, United States District Court Northern District Of California San Francisco Division, Case 3:23-cv-00201, disponible en <https://stablediffusionlitigation.com/pdf/00201/1-1-stable-diffusion-complaint.pdf>, consultado el 20 de enero de 2023.

49 GETTY IMAGES (US), Inc. v. STABILITY AI, Inc. Demand for Jury Trial, United States District Court for The District of Delaware, Case 1:23-cv-00135-UNA, disponible en <https://www.courtlistener.com/docket/66788385/getty-images-us-inc-v-stability-ai-inc/>, consultado el 15 de febrero de 2023.

50 Por ejemplo, en la demanda de Getty Images se hace alusión al uso por parte de las creaciones de los sistemas de IA de versiones modificadas de la marca de agua de Getty Images, cuyo uso no solo es de baja calidad, poco atractivo u ofensivo, sino también que crea confusión en relación con el origen de las imágenes y provoca riesgo de asociación que no existe. Lo cual no solo vulnera las leyes de protección de marcas, sino también de competencia. GETTY IMAGES (US), Inc. v. STABILITY AI, Inc. Demand for Jury Trial, United States District Court for The District of Delaware, Case 1:23-cv-00135-UNA, disponible en <https://www.courtlistener.com/docket/66788385/getty-images-us-inc-v-stability-ai-inc/>, consultado el 15 de febrero de 2023.

51 SARAH ANDERSEN, an individual; KELLY MCKERNAN, an individual; KARLA ORTIZ, an individual, Individual and Representative Plaintiffs, v. STABILITY AI LTD., a UK corporation; STABILITY AI, INC., a Delaware corporation; MIDJOURNEY, INC., a Delaware corporation; DEVIANTART INC, a Delaware corporation, Class Action Complaint, United States District Court Northern District Of California San Francisco Division, Case 3:23-cv-00201, disponible en <https://stablediffusionlitigation.com/pdf/00201/1-1-stable-diffusion-complaint.pdf>, consultado el 20 de enero de 2023, Sección 204 y 206.

tución que pudiera tener lugar como fuente de los datos y que ocurre precisamente a partir de la obtención de datos sintéticos. El impacto de la piratería de la IA también queda corroborado en la capacidad que estos modelos tienen para crear datos de entrenamiento sintéticos. Lo cual significa también el uso de menor cantidad de datos obtenidos a partir de obras reales. Todo lo cual afecta no solo en materia de sustitución, sino también, el propio desarrollo de la industria. Los datos sintéticos constituyen una valiosa herramienta para proteger la privacidad de los datos así como los derechos de los autores de las obras que se utilizan en los sistemas de IA generativos. Sin embargo, debe considerarse que el dato sintético es generado precisamente a partir de “datos reales”, sobre los que pueden existir otros derechos como son los derechos de autor y a la privacidad.

Empero lo más importante a tener en cuenta es que el dato sintético es un sustituto de las obras, creados a partir de estas. De hecho, se calcula que en el 2024 el 60% de los datos utilizados en el desarrollo de la IA serán sintéticos<sup>52</sup>. Lo cual tiene un impacto en la cadena de valor de los datos y, en particular, de aquellos que son extraídos a partir de obras y demás prestaciones protegidas por los derechos de autor. Si bien, por un lado, el dato sintético provocará una sustitución en el uso de aquellos datos que son extraídos de obras y demás prestaciones, protegiéndolos; por otro lado, es lógico que incrementará el valor de estos. Ello obedece al hecho de que para que el dato sintético se parezca cada vez más a los datos reales tendrá que utilizar datos reales de calidad, cuyo acceso debería ser más costoso, siendo los datos sintéticos los que tendrían mayores usos en el ámbito de la innovación y la industria de la IA.

Por ende, lo que los sistemas de IA generativos en la actualidad están haciendo no es solo desarrollar modelos de negocios que sustituyen el valor de las obras en el mercado, sino además, el valor de las obras como fuente de los datos. El impacto del uso de obras en el entrenamiento de los sistemas de IA no es solo en materia de mercado de las industrias creativas, sino también, del propio mercado de datos. De ahí que, no reconocer la necesidad de la autorización para su utilización en esta actividad significaría desaprovechar además una apuesta de valor que las obras tienen en sí mismas como fuente de los datos, utilizadas en el entrenamiento de los modelos de IA. Valor del cual no se están beneficiando los creadores sino, solo los utilizadores, lo que incide, de forma determinante, en la brecha de valor de los derechos de autor.

## CONCLUSIONES

Analizar el complejo fenómeno de la piratería y su relación con los sistemas de IA generativos es un reto, que excede, por mucho, los límites de este trabajo. El esfuerzo por

realizar conclusiones al respecto debe partir de reconocer las propias limitaciones que existen en la temática. La primera y más importante es que es un fenómeno en constante y vertiginoso desenvolvimiento. Ello implica que no conocemos a ciencia cierta cuál es su verdadero impacto en la actualidad, así como desconocemos cuánto avanza la tecnología de la IA en esta materia.

Sin embargo, a partir de la información a la que se ha tenido acceso y del análisis realizado se puede reconocer que ciertamente cuando se utilizan obras protegidas por los derechos de autor para entrenar sistemas de IA generativos, estamos en presencia de una infracción de estos derechos. El uso que actualmente se realiza sin recabar la autorización de los titulares debe ser considerado pirata que provoca perjuicios no solo desde el ámbito legal, al infringir normas de derechos de autor, sino también, un perjuicio económico que se agrava con una situación de competencia desleal y desigualdad tecnológica y económica. Concebir que existe una infracción y, particularmente, un acto de piratería debería fundamentar la necesidad de reconocer la participación de los autores y demás titulares en los beneficios que se obtienen a partir del uso de sus obras y demás prestaciones artísticas, al propio tiempo que le brinda la posibilidad de acudir a instrumentos nacionales e internacionales de observancia de los derechos de autor.

La perspectiva que se sostiene no parte de prohibir dicho uso, sino que este tiene que ser realizado sin perjuicios para los autores y/o creadores, es decir, sin que se excluyan a estos de los beneficios que aporta dicha utilización. Excluir de facto a los autores de participar de estos beneficios agudiza la brecha de valor que existe, así como las condiciones de precariedad de los creadores. En un contexto, en el cual el desarrollo de la tecnología de IA ya de por sí representa un importante reto en materia de creatividad, competencia y cadenas de valor, se le debe sumar la posibilidad del aprovechamiento económico de un uso que demuestra ser redituable.

La forma en la que se determine como puede ser dicha participación y a partir de cuáles mecanismos jurídicos, dígame licencias, remuneraciones obligatorias, etc., ya es análisis de otras propuestas de lege ferenda. En dichas propuestas se deberá de tener en cuenta los costos de transacción, su impacto en la innovación y el desarrollo de la economía digital, así como la participación de otros actores como podrían ser las entidades de gestión colectiva. Sin embargo, ello merece un análisis más detallado y profundo.

## REFERENCIAS BIBLIOGRÁFICAS

BLACKBURN, DAVID; EISENACH, JEFFREY A.; HARRISON, DAVID JR. (2019): *Impacts Of Digital Video Piracy On The U.S. Economy*, Chamber's

52 Toews (2022).

- Global Innovation Policy Center, U.S. Chamber of Commerce, NERA Economic Consulting, Washington, disponible en <https://www.theglobalipcenter.com/wp-content/uploads/2019/06/Digital-Video-Piracy.pdf>, consultado el 15 de enero de 2023.
- CARLINI, NICHOLAS; HAYES, JAMIE; NASR, MILAD; JAGIELSKI, MATTHEW; SEHWAG, VIKASH; TRAMÈR, FLORIAN; BALLE, BORJA; IPPOLITO, DAPHNE; WALLACE, ERIC (2023): "Extracting Training Data from Diffusion Models", disponible en <https://arxiv.org/abs/2301.13188>, consultado el 15 de febrero de 2023, págs. 1-31.
- CARR, ANDREW (2022): "Diffusion models for document synthesis", disponible en <https://gretel.ai/blog/diffusion-models-for-document-synthesis>, consultado el 25 de febrero de 2023.
- CARROLL, MICHAEL W. (2019): "Copyright and the Progress of Science: Why Text and Data Mining Is Lawful", UC Davis Law Review, University of California, Vol. 53, págs. 893-964.
- DALUGDUG, MANDY (2022): "Record Industry Clamps Down on Ai-Based Music Extractors that Infringe on Copyrights", disponible en <https://www.musicbusinessworldwide.com/record-industry-clamps-down-on-ai-based-music-extractors-that-infringe-on-copyrights/>, consultado el 15 de enero de 2023.
- DANAHER, BRETT; SMITH, MICHAEL D.; TELANG, RAHUL (2020): "Piracy Landscape Study: Analysis of Existing and Emerging Research Relevant to Intellectual Property Rights (IPR) Enforcement of Commercial-Scale Piracy", USPTO Economic Working Paper, No. 2020-02, April, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3577670](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3577670).
- FREITAS, DENIS DE (1992): "Pirateria de la Propiedad Intelectual y medidas necesarias para combatirla", Boletín de Derecho de Autor de la UNESCO, volumen XXVI, No. 3, págs. 1-19.
- GUADAMUZ, ANDRES, (2023): "A Scanner Darkly: Copyright Infringement in Artificial Intelligence Inputs and Outputs", disponible en SSRN: <https://ssrn.com/abstract=4371204> o <http://dx.doi.org/10.2139/ssrn.4371204>, consultado 3 de marzo de 2023.
- HEIKKILÄ, MELISSA (2023): "Los modelos de IA reproducen fotos de personas reales e imágenes con derechos de autor", MIT Technology Review, traducido por Ana Milutinovic, 10 de febrero, disponible en <https://www.technologyreview.es/s/15072/los-modelos-de-ia-reproducen-fotos-de-personas-reales-e-imagenes-con-derechos-de-autor>, consultado el 15 de febrero de 2023.
- HEIKKILÄ, MELISSA (2022): "El artista que inspira miles de obras generadas por IA: "Empieza a parecer una amenaza", MIT Technology Review, traducido por Ana Milutinovic disponible en <https://www.technologyreview.es/s/14587/el-artista-que-inspira-miles-de-obras-generadas-por-ia-empieza-parecer-una-amenaza>, consultado el 15 de febrero de 2023.
- JAHANIAN, ALI; PUIG, XAVIER; TIAN, YONGLONG; ISOLA, PHILLIP (2022): "Generative models as a data source for Multiview representation Learning", disponible en <https://arxiv.org/pdf/2106.05258.pdf>, consultado el 25 de febrero de 2023, págs.1-22.
- PANETHIERE, DARRELL (2005): "La persistencia de la piratería y sus consecuencias para la creatividad, la cultura y el desarrollo sostenible", estudio elaborado a petición de la Secretaría de la UNESCO para la 13a reunión del Comité Intergubernamental de Derecho de Autor. GC (1971)/XIII/6, París, 31 de mayo, disponible en [https://unesdoc.unesco.org/ark:/48223/pf0000139651\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000139651_spa), consultado el 10 de marzo de 2023, págs. 1-20.
- STASSEN, MURRAY (2022): "Over 1,000 Songs with Human-Mimicking AI Vocals Have Been Released by Tencent Music in China. One Of Them Has 100m Streams", disponible en <https://www.musicbusinessworldwide.com/over-1000-songs-human-mimicking-ai-vocals-have-been-released-by-tencent-music-in-china-one-of-them-has-over-100m-streams>, consultado el 18 de enero de 2023.
- TOEWS, ROB (2022): "Synthetic Data Is About To Transform Artificial Intelligence Contributor", Forbes, disponible en <https://www.forbes.com/sites/robtoews/2022/06/12/synthetic-data-is-about-to-transform-artificial-intelligence/?sh=3fc099247523>, consultado el 15 de marzo de 2023.
- VINCENT, JAMES (2022): 'An engine for the imagination': the rise of AI image generators An interview with Midjourney founder David Holz", Aug 2, disponible en <https://www.theverge.com/2022/8/2/23287173/ai-image-generation-art-midjourney-multiverse-interview-david-holz>, consultado el 5 de marzo de 2023.