

Datasets de entrenamiento en la inteligencia artificial generativa: desafíos para el derecho de autor

Training datasets in generative artificial intelligence: challenges for copyright

César Ricardo Castillo Velazco¹

Autor:

¹ Centro de Ingeniería y Desarrollo Industrial (CIDESI).
cesar.castillo@cidesi.edu.mx
<https://orcid.org/0000-0002-1130-1816>

Recibido: 25/DICIEMBRE/2025

Aprobado: 22/MAYO/2026

Publicación online: 02/JUNIO/2026

Cómo citar/How to cite:

Castillo Velazco, C. (2026). *Datasets de entrenamiento en la inteligencia artificial generativa: desafíos para el derecho de autor*. *Revista Especializada en Investigación Jurídica*, 10(19), 1-17.
<https://doi.org/10.20983/reij.2026.2.1>

Licencia:

Este trabajo es de acceso abierto distribuido bajo los términos y condiciones de la Licencia Creative Commons Attribution ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/))



© 2026 Autor1

RESUMEN

La inteligencia artificial generativa (IAG) ha transformado de manera significativa los procesos de creación y producción de contenidos, al basarse en el entrenamiento de modelos con grandes volúmenes de datos que, en numerosos casos, incluyen obras protegidas por derecho de autor. El *objetivo* de esta investigación es analizar las implicaciones jurídicas derivadas del uso de obras protegidas por derecho de autor en los *datasets* de entrenamiento de la IAG, con el fin de identificar las principales tensiones normativas y brechas regulatorias existentes. La metodología es de carácter cualitativo y se basa en una revisión doctrinal y normativa, complementada con una perspectiva de derecho comparado. El análisis comparado realizado permitió identificar que el uso masivo de obras protegidas en procesos de entrenamiento implica actos jurídicamente relevantes de reproducción y transformación que no encuentran una respuesta clara en el marco jurídico internacional vigente. De igual manera, la opacidad técnica de los modelos generativos dificulta la trazabilidad de las obras utilizadas y limita el ejercicio efectivo de los derechos de autor. Como conclusión, se sostiene la necesidad de repensar los fundamentos tradicionales del derecho de autor para avanzar hacia un régimen jurídico más flexible y equilibrado, que incorpore esquemas proporcionales de compensación, sin obstaculizar el progreso tecnológico.

Palabras clave: *datasets* de entrenamiento; derecho de autor; inteligencia artificial generativa.

ABSTRACT

Generative artificial intelligence (GAI) has significantly transformed content creation and production processes by training models on large volumes of data that, in many cases, include works protected by copyright. The aim of this research is to analyze the legal implications of using copyrighted works in GAI training datasets and to identify the main normative tensions and existing regulatory gaps. The methodology is qualitative and based on a doctrinal and normative review, complemented by a comparative law perspective. The comparative analysis conducted made it possible to identify that the large-scale use of copyrighted works in training processes entails legally relevant acts of reproduction and transformation that currently lack a clear response within the international legal framework.

Likewise, the technical opacity of generative models hinders the traceability of the works used and limits the effective exercise of copyright protection. In conclusion, this study argues for rethinking the traditional foundations of copyright law to move toward a more flexible and balanced legal regime that incorporates proportional compensation mechanisms without hindering technological progress.

Keywords: copyright; generative artificial intelligence; training datasets.

INTRODUCCIÓN

El desarrollo reciente de la inteligencia artificial generativa (IAG) ha transformado de manera significativa los procesos de creación y producción de contenidos, al permitir la generación automatizada de textos, imágenes, música, entre otros, mediante el entrenamiento de modelos con grandes volúmenes de datos. Estos modelos, particularmente los denominados modelos fundacionales, dependen de conjuntos de datos (*datasets*) masivos y heterogéneos, que con frecuencia incorporan obras protegidas por derecho de autor sin autorización previa de sus autores o titulares (Bommasani *et al.*, 2021; Henderson *et al.*, 2023). Esta práctica ha colocado a la IAG en el centro de un debate jurídico contemporáneo sobre la adecuación del régimen autoral frente a los avances tecnológicos.

Desde la teoría clásica del derecho de autor, la protección se sustenta en la creatividad humana, la originalidad de la obra y la atribución de derechos exclusivos que permiten controlar su reproducción y transformación. No obstante, la utilización automatizada —y a gran escala— de obras en los procesos de entrenamiento de la IAG desafía estos fundamentos, al tratarse de usos técnicamente necesarios, opacos, difícilmente rastreables, que no encajan plenamente en las categorías tradicionales del derecho de autor (Gervais, 2020). La literatura reciente ha advertido que el entrenamiento de modelos generativos implica actos jurídicamente relevantes de reproducción y transformación, mientras que la opacidad algorítmica limita la trazabilidad de las obras y debilita la tutela efectiva de los derechos de autor (United States Copyright Office, 2025; Kretschmer, Margoni y Oruç, 2024; Buick, 2024; Pasetti *et al.*, 2025; Lucchi, 2025).

En el ámbito normativo internacional, las respuestas regulatorias no han sido homogéneas. Aunque algunas jurisdicciones han comenzado a introducir excepciones específicas para la minería de textos y datos, otros ordenamientos carecen de reglas claras sobre el uso de obras protegidas en los *datasets* de entrenamiento, lo que genera incertidumbre jurídica para creadores y desarrolladores tecnológicos. En este contexto, el *objetivo* general del presente artículo es analizar las implicaciones jurídicas derivadas del uso de obras protegidas por derecho de autor en los *datasets* de entrenamiento de la IAG, a fin de identificar tensiones normativas y brechas regulatorias, así como explorar alternativas que permitan conciliar la innovación tecnológica con la protección efectiva de los derechos de autor.

Para alcanzar dicho objetivo, se adopta un enfoque cualitativo comparado, basado en la revisión doctrinal y normativa, orientado al análisis de las principales tensiones regulatorias derivadas del uso de obras protegidas en los *datasets* de entrenamiento. El artículo se estructura en apartados dedicados al estudio de la naturaleza de los *datasets*, los fundamentos autorales aplicables, las implicaciones jurídicas del entrenamiento, la protección del contenido generado,

el análisis comparado de distintos modelos regulatorios y la exploración de posibles alternativas normativas.

METODOLOGÍA

La presente investigación adopta un enfoque cualitativo de carácter jurídico-doctrinal, el cual se basa en el análisis de fuentes normativas, doctrinales e institucionales relacionadas con la IAG y el derecho de autor. El estudio se desarrolla desde una perspectiva comparada, orientada a identificar las principales tensiones jurídicas derivadas del uso de obras protegidas en los *datasets*. La metodología se estructuró en torno a tres ejes: a) la naturaleza jurídica de los *datasets* y de las operaciones técnicas para el entrenamiento de la IAG; b) la posible afectación a los derechos patrimoniales y morales del autor; y c) las respuestas regulatorias adoptadas en distintas jurisdicciones frente a estos desafíos tecnológicos.

En este sentido, se analizaron los modelos regulatorios de Estados Unidos, la Unión Europea y China, que constituyen referentes internacionales diferenciados en materia de gobernanza de la inteligencia artificial (IA), contrastándolos con el marco jurídico mexicano. Las fuentes utilizadas incluyeron: literatura especializada reciente, instrumentos internacionales, legislación y documentos institucionales relevantes. Esta investigación adopta un enfoque crítico y prospectivo, orientado a evaluar la suficiencia del régimen tradicional del derecho de autor frente al entrenamiento de modelos generativos, así como a explorar posibles alternativas regulatorias.

NATURALEZA Y FUNCIÓN DE LOS DATASETS DE ENTRENAMIENTO

Para comprender la relevancia jurídica de los *datasets* de entrenamiento en la IAG, resulta necesario partir de una aproximación conceptual a la IA. Si bien no existe una definición única, la IA suele entenderse como el conjunto de técnicas computacionales orientadas al diseño y entrenamiento de sistemas capaces de ejecutar tareas que tradicionalmente requieren inteligencia humana, tales como el reconocimiento de patrones, el aprendizaje, la resolución de problemas y la toma de decisiones. La literatura contemporánea ha destacado que estas definiciones pueden agruparse en función de si conciben a la IA como sistemas que imitan el comportamiento humano o sistemas orientados a actuar de manera racional, optimizando objetivos predefinidos (Russell y Norvig, 2021).

Dentro de este campo, la IAG representa una de las ramas más disruptivas, al caracterizarse por la capacidad de generar nuevos contenidos (texto, imágenes, música, video o código) mediante el uso de modelos generativos entrenados con grandes volúmenes de datos. A diferencia de aplicaciones tradicionales de IA centradas en tareas de clasificación o predicción, los sistemas generativos producen resultados novedosos a partir de la identificación de patrones y estructuras subyacentes en los datos de entrenamiento, lo que introduce un componente de producción automatizada de contenidos, que ha intensificado los debates técnicos, sociales y jurídicos en torno a su funcionamiento y efectos (Bommasani *et al.*, 2021; Sengar *et al.*, 2024).

El funcionamiento de la IAG, se sustenta principalmente en técnicas de aprendizaje automático profundo (*deep learning*), basadas en redes neuronales artificiales de gran escala. Para el diseño de estos modelos, se requiere de una fase de entrenamiento, en la cual se ajustan de manera iterativa millones de parámetros internos a partir de los datos suministrados, aprendiendo regularidades estadísticas presentes en el lenguaje, las imágenes o el sonido. El desempeño posterior del modelo depende directamente de esta etapa, lo que explica la importancia crítica del acceso a grandes volúmenes de datos diversos y de alta calidad (Bommasani *et al.*, 2021).

Estos desarrollos han dado lugar a los denominados modelos fundacionales, entendidos como modelos entrenados sobre *datasets* masivos, heterogéneos y de propósito general, capaces de transferir lo aprendido a múltiples tareas mediante procesos de afinamiento (*fine-tuning*). Estos modelos concentran capacidades de representación, razonamiento y generación en una sola arquitectura, lo que los convierte en infraestructura básica para una amplia gama de aplicaciones. No obstante, este diseño escalable también ha generado preocupaciones relativas a la concentración de poder tecnológico y el control de los datos en manos de un número reducido de actores con capacidad para asumir los costos computacionales y de recopilación de datos que estos modelos requieren (Bommasani *et al.*, 2021; Comisión Europea, 2022).

En este contexto, los *datasets* de entrenamiento no constituyen meros repositorios de información, sino el insumo central que determina la versatilidad, el rendimiento y la capacidad generativa de los modelos. La diversidad y procedencia de los datos influyen directamente en la calidad de los resultados, pero también incrementan los riesgos asociados a la incorporación de sesgos, la reproducción de estereotipos y la utilización de contenidos protegidos por derechos de autor sin autorización. Esta dependencia estructural de datos masivos explica por qué el debate sobre los *datasets* de entrenamiento trasciende el ámbito técnico y adquiere una dimensión jurídica relevante, vinculada con la autoría, el uso legítimo de obras protegidas y la necesidad de mecanismos de licenciamiento y/o compensación (Henderson *et al.*, 2023; Gervais, 2020).

Los conjuntos de datos utilizados para entrenar modelos generativos pueden conformarse a partir de datos abiertos o disponibles públicamente en internet, así como de datos privados recopilados y procesados por empresas especializadas. Estos conjuntos suelen incluir textos literarios, artículos científicos y periodísticos, imágenes, fotografías, música, etcétera, muchos de los cuales se encuentran protegidos por derechos de autor. Esta práctica amplía el potencial creativo de los sistemas de IAG, pero al mismo tiempo incrementa el riesgo de afectaciones a los derechos de los autores y titulares, especialmente cuando no existen mecanismos claros de autorización, trazabilidad o compensación (Henderson *et al.*, 2023; United States Copyright Office, 2025).

Finalmente, el entrenamiento de modelos generativos enfrenta un problema adicional relativo a la opacidad algorítmica, comúnmente descrita como el fenómeno de la “caja negra”. La complejidad de los modelos y de los procesos de aprendizaje dificulta identificar qué datos específicos influyeron en la generación de un resultado concreto, lo que impide conocer con precisión qué obras fueron utilizadas y en qué medida (Mitchell, 2025). Esta falta de trazabilidad no solo genera incertidumbre técnica, sino que plantea interrogantes jurídicos relevantes en múltiples materias, específicamente en el campo del derecho de autor, al dificultar la supervisión, la rendición de cuentas y el ejercicio efectivo de los derechos por parte de los titulares afectados.

La relevancia técnica de los *datasets* de entrenamiento trasciende el ámbito técnico y plantea interrogantes jurídicos de primer orden, al involucrar la incorporación masiva de contenidos que, en muchos casos, se encuentran protegidos por el derecho de autor. Esta circunstancia hace necesario analizar cómo los fundamentos tradicionales del régimen autoral, se aplican a los procesos de integración y uso de obras en dichos conjuntos de datos.

FUNDAMENTOS DEL DERECHO DE AUTOR APLICABLES AL USO DE OBRAS EN DATASETS DE ENTRENAMIENTO

Cuando una obra es incorporada en un conjunto de datos para el desarrollo o afinamiento de un modelo de IAG, no puede ser considerada como un dato neutral o carente de relevancia jurídica. Por el contrario, se trata del uso de una creación intelectual protegida por el régimen jurídico

del derecho de autor, cuya incorporación implica actos de relevancia jurídica. La noción clásica de obra como “expresión personal de la inteligencia que se manifiesta bajo una forma perceptible, con suficiente originalidad y apta para ser difundida o reproducida” continúa siendo un punto de partida válido para el análisis contemporáneo (Lipszyc, 1993, p. 48), pero debe ser reinterpretada a la luz de los entornos digitales y de los procesos automatizados de tratamiento masivo de información.

Lo anterior implica que para acceder a esta protección no es necesario cumplir mayores requisitos que los esenciales. La protección autoral surge de manera automática desde el momento en que la obra es creada y fijada en un soporte material, sin necesidad de registro ni formalidades adicionales. No se exige novedad en sentido estricto, mérito artístico ni finalidad específica, bastando con que la obra refleje un mínimo de originalidad derivado de la impronta personal del autor. Estos principios, consolidados en la doctrina autoral, mantienen plena vigencia en el entorno digital y resultan especialmente relevantes frente a prácticas tecnológicas que tienden a diluir la individualidad de las obras al incorporarlas en grandes repositorios de datos (De la Parra, 2015).

En cuanto a los sujetos del derecho de autor, es posible identificar dos figuras diferentes: el autor y el titular. El sistema autoral continental continúa reconociendo al autor exclusivamente como persona física, en tanto único sujeto capaz de realizar actos creativos. Esta concepción encuentra sustento en los instrumentos internacionales. Particularmente, el artículo 6 Bis del Convenio de Berna (1886) hace referencia a ciertos derechos que se mantendrán, incluso después de la muerte del autor, lo que permite concluir que este acuerdo reconoce el carácter de autor a una persona física. Las personas jurídicas pueden ostentar únicamente una titularidad derivada, ya sea por cesión, presunción legal o transmisión *mortis causa*, sin que ello les otorgue la condición de autores originarios (Lipszyc, 1993). En los países con sistema *copyright* existen diversos supuestos en los que se reconoce la calidad de autor a las personas jurídicas, ya sea que se trate de obras por encargo, derivadas de una relación laboral o las producciones cinematográficas (Ginsburg y Budiardjo, 2019).

El derecho de autor confiere a los creadores un conjunto de facultades que se dividen tradicionalmente en derechos morales y patrimoniales. Mientras los primeros protegen el vínculo personal del autor con su obra (paternidad, integridad y divulgación), los segundos le permiten autorizar o prohibir actos de explotación económica, tales como reproducción, distribución, comunicación pública y transformación. En el contexto de los *datasets* de entrenamiento, estas facultades adquieren una relevancia central, pues la incorporación de obras protegidas suele implicar copias digitales, fijación en servidores, indexación y, en ocasiones, adaptación o fragmentación del contenido, actos que encajan en las categorías clásicas de reproducción y transformación (United States Copyright Office, 2025).

En las últimas décadas, las tecnologías digitales han ido perfeccionándose y se ha llevado a su campo la creación intelectual, el uso y la explotación de las obras. En lo que respecta al uso de una obra en medios digitales, al igual que como acontece en el mundo físico, se requiere de la autorización de su autor o titular, tal y como lo establece el artículo 9 del Convenio de Berna (1886), al señalar que aquellos gozarán “del derecho exclusivo de autorizar la reproducción de sus obras por cualquier procedimiento y bajo cualquier forma”. En el ámbito europeo, esta postura fue confirmada por la jurisprudencia del Tribunal de Justicia de la Unión Europea (TJUE), particularmente en el caso de *Infopaq International A/S v. Danske Dagblades Forening* (C-5/08, 2009), en el cual se determinó que la reproducción de fragmentos digitales de obras constituye una reproducción parcial y, por tanto, requiere de autorización. Posteriormente, la Directiva 2001/29/CE (InfoSoc) y la más reciente Directiva 2019/790/UE consolidaron la idea de que toda copia digital, ya sea temporal o permanente, implica una reproducción. Lo anterior resulta especialmente significativo para el análisis del entrenamiento

de modelos generativos, en tanto los procesos de minería de datos y aprendizaje automático requieren necesariamente la realización de copias técnicas de las obras utilizadas (Comisión Europea, 2022).

Frente a estos derechos exclusivos, los ordenamientos jurídicos contemplan limitaciones y excepciones orientadas a equilibrar la protección autoral con intereses públicos, como el acceso a la cultura, la educación y la investigación. Su fundamento se encuentra en el equilibrio entre la protección de los derechos exclusivos del autor y el interés público de garantizar el acceso a la cultura, la educación, la información, la investigación, entre otros. Se trata de expresiones del Principio de Proporcionalidad, el cual implica que la protección autoral no puede reconocerse de manera absoluta si con ello se restringen de forma injustificada otros derechos fundamentales (De la Parra, 2015).

Estos supuestos se fundamentan en la regla de los tres pasos, introducida en el inciso 2) del artículo 9 del Convenio de Berna (1886), que exige que los usos permitidos sean especiales, no interfieran con la explotación normal de la obra y no causen un perjuicio injustificado a los intereses legítimos del autor. En el entorno digital, las limitaciones y excepciones han cobrado una nueva dimensión, particularmente en relación con la minería de textos y datos, donde algunos ordenamientos han optado por introducir excepciones específicas acompañadas de salvaguardas y mecanismos de exclusión (Comisión Europea, 2022).

Los avances de la Unión Europea en esta materia constituyen un referente relevante, en la medida en que han impulsado la modernización de su marco normativo mediante la introducción de excepciones específicas orientadas a equilibrar la protección del derecho de autor con la innovación tecnológica y el interés público. Por su parte, en Estados Unidos, si bien no existe una excepción estatutaria expresa para la minería de textos y datos o el entrenamiento de modelos de IAG, la doctrina del *fair use* ha operado como un mecanismo flexible, que permite evaluar, caso por caso, la licitud de estos usos, a partir de criterios como el propósito transformativo y el impacto en el mercado de las obras. En contraste, México, al igual que otros ordenamientos jurídicos, carece de disposiciones específicas que regulen el uso de obras protegidas en los *datasets* de entrenamiento, lo que genera un alto grado de incertidumbre y limita los mecanismos de protección efectiva de los autores, además de plantear el riesgo de incumplimiento de obligaciones internacionales asumidas en materia de derecho de autor.

Lo anterior pone de manifiesto la tensión entre innovación tecnológica y la protección autoral, además de revelar que el marco normativo actual resulta insuficiente para atender la magnitud del fenómeno estudiado. Los fundamentos del derecho de autor fueron concebidos para contextos de uso y explotación de las obras, inicialmente en el mundo físico y luego se extendieron a lo digital, pero de manera individualizada, no para su incorporación masiva en *datasets* que alimentan modelos capaces de generar miles o millones de resultados en cuestión de segundos. La persistencia de esta situación implica un doble riesgo: por un lado, la desprotección efectiva de los derechos morales y patrimoniales de los autores; y por otro, el freno a la innovación tecnológica derivado de la inseguridad jurídica. De ahí que la doctrina reciente haya comenzado a explorar alternativas regulatorias intermedias, como licencias colectivas, mecanismos de compensación o regalías, y soluciones tecnológicas de trazabilidad, orientadas a conciliar el desarrollo de la IAG con la tutela efectiva del derecho de autor y otros derechos fundamentales (Samuelson, 2023).

Los fundamentos puntualizados permiten advertir que la incorporación de obras protegidas en los *datasets* de entrenamiento de la IAG no constituye una práctica jurídicamente neutra, sino un conjunto de actos con relevancia directa para los derechos exclusivos de los autores y titulares. A partir de esta base resulta necesario analizar de manera específica las implicaciones jurídicas que se derivan de dichos procesos de entrenamiento, particularmente en lo relativo a

la posible configuración de infracciones, la afectación a los derechos morales y patrimoniales, así como a las dificultades prácticas para su tutela efectiva en entornos caracterizados por la automatización y la opacidad tecnológica.

IMPLICACIONES JURÍDICAS DEL USO DE OBRAS PROTEGIDAS EN LOS PROCESOS DE ENTRENAMIENTO DE LA IAG

Como se ha señalado, la incorporación de obras protegidas por derecho de autor en los *datasets* de entrenamiento constituye un uso jurídicamente relevante de creaciones intelectuales. El régimen autoral parte de la premisa de que toda utilización de una obra, con independencia del medio o de la tecnología empleada, se encuentra comprendida dentro de los derechos patrimoniales exclusivos del autor o del titular. En consecuencia, actos como la digitalización, el almacenamiento, la reproducción, total o parcial, y la posterior utilización de las obras para el entrenamiento de sistemas de IAG no quedan excluidos del ámbito de protección autoral por el solo hecho de ser ejecutados mediante procesos automatizados o algorítmicos (United States Copyright Office, 2025).

Desde una perspectiva técnica, la construcción de *datasets* de entrenamiento implica necesariamente la realización de copias digitales de las obras seleccionadas, su fijación en soportes electrónicos y su indexación dentro de repositorios susceptibles de tratamiento computacional. Tales operaciones encajan en la noción de reproducción prevista en el artículo 9 del Convenio de Berna, en la medida en que suponen la creación de copias perceptibles mediante medios tecnológicos, incluso cuando estas sean de carácter temporal o parcial. Asimismo, la preparación de los *datasets* puede requerir la transformación de las obras originales, por ejemplo, mediante la fragmentación de textos, la redimensión de imágenes o la modificación de pistas sonoras, lo que puede dar lugar a la generación de obras derivadas sin autorización del titular legítimo (Comisión Europea, 2022).

De este modo, el entrenamiento de la IAG no se limita a la mera conservación de contenidos, sino que involucra una combinación de actos de reproducción, transformación y, en ciertos casos, comunicación pública, particularmente cuando los *datasets* son compartidos, comercializados o puestos a disposición de terceros. Desde el punto de vista autoral, cada uno de estos actos requiere, en principio, la autorización del autor o titular, lo que convierte estas prácticas en potenciales supuestos de infracción cuando se realizan sin respaldo jurídico suficiente (United States Copyright Office, 2025; Comisión Europea, 2025).

La afectación a los derechos patrimoniales resulta especialmente evidente, debido a que las obras incorporadas en los procesos de entrenamiento contribuyen al desarrollo de sistemas que generan un valor económico significativo para desarrolladores y empresas tecnológicas, sin que los autores reciban compensación ni conserven un control efectivo sobre la explotación de sus creaciones. Esta problemática se intensifica cuando los modelos generativos producen resultados que, aun sin ser idénticos, presentan un alto grado de similitud con las obras utilizadas como insumo, lo que plantea interrogantes no resueltas sobre la posible configuración de reproducciones indebidas, plagio o creación de obras derivadas en el contexto de la generación automatizada de contenidos (Samuelson, 2023; Gervais, 2020).

Por su parte, los derechos morales también se ven comprometidos, toda vez que la utilización de obras en *datasets* de entrenamiento puede diluir el vínculo personal entre el autor y su creación, afectando el derecho de paternidad, en la medida en que los resultados generados por la IAG no reconocen ni atribuyen la autoría de las obras que influyeron directa o indirectamente en su producción. Asimismo, la fragmentación o modificación de las obras durante el proceso de entrenamiento puede vulnerar el derecho a la integridad, al alterarse la obra sin consideración a la voluntad del autor (Ginsburg y Budiardjo, 2019).

Estas afectaciones se ven agravadas por la ya referida opacidad técnica que caracteriza a los modelos generativos. La complejidad de los procesos de entrenamiento y de generación de resultados dificulta, incluso para los propios desarrolladores, identificar qué datos fueron determinantes en la obtención de un contenido específico. Esta falta de trazabilidad limita de manera significativa la posibilidad de que los autores ejerzan acciones jurídicas para la tutela de sus derechos, ya que resulta extremadamente complejo demostrar la relación causal entre una obra concreta y un resultado generado por la IAG. Cuando se considera, además, que los *datasets* pueden contener millones de registros, la carga de verificar si una obra fue utilizada se vuelve prácticamente inasumible para los titulares de derechos (Lucchi, 2025).

Este escenario genera un desequilibrio estructural entre los actores involucrados, incrementa el riesgo de infracciones masivas y sistemáticas, y puede erosionar los incentivos para la creación intelectual. En consecuencia, la literatura reciente ha subrayado la necesidad de diseñar mecanismos jurídicos y tecnológicos que permitan equilibrar el fomento a la innovación con la salvaguarda efectiva de los derechos de autor, evitando que la opacidad algorítmica se traduzca en desprotección jurídica y debilitamiento del sistema autoral en su conjunto (Samuelson, 2023; Lucchi, 2025).

Las implicaciones jurídicas derivadas del uso de obras protegidas en los procesos de entrenamiento de la IAG no se agotan en la fase de incorporación y tratamiento de los *datasets*, sino que proyectan sus efectos sobre la naturaleza y el estatus jurídico de los resultados producidos por estos sistemas. En particular, la posible afectación a los derechos patrimoniales y morales de los autores, así como las dificultades de trazabilidad y atribución, plantean interrogantes adicionales respecto a si el contenido generado puede considerarse jurídicamente protegido, bajo qué criterios y a quién correspondería, en su caso, la titularidad de los derechos derivados. Estas cuestiones serán abordadas brevemente en el apartado siguiente.

NATURALEZA JURÍDICA Y PROTECCIÓN DEL CONTENIDO GENERADO POR LA IAG

La determinación de la naturaleza jurídica del contenido generado por la IAG constituye uno de los debates más relevantes del derecho de autor contemporáneo. La primera cuestión por resolver es si los resultados producidos por sistemas de IAG pueden considerarse creaciones protegidas en el sentido jurídico tradicional. La posición mayoritaria, tanto en la doctrina como en la práctica administrativa comparada, sostiene que el derecho de autor presupone la existencia de una intervención humana creativa capaz de imprimir una impronta personal y un grado mínimo de originalidad. En ausencia de dicha intervención humana sustantiva, los productos generados de manera autónoma por sistemas de IAG carecen de autoría y, por ende, no pueden ser calificados como obras protegidas por el régimen autoral (United States Copyright Office, 2023; Lucchi, 2025; Gervais, 2020).

Esta interpretación ha sido reiterada por diversas autoridades, que han enfatizado que los sistemas de IA no pueden ser considerados autores ni tampoco los desarrolladores o usuarios pueden atribuirse automáticamente la autoría de los resultados generados, sin demostrar una contribución creativa relevante. De este modo, se reafirma el principio del derecho de autor, según el cual la protección se justifica como reconocimiento de la creatividad humana y no como recompensa a procesos puramente automatizados (Ginsburg y Budiardjo, 2019).

Un segundo nivel de análisis consiste en examinar si los contenidos generados por la IAG pueden calificarse como obras derivadas, en la medida en que los modelos generativos se entrenan a partir de *datasets* que incluyen obras protegidas. Desde una perspectiva jurídica, la obra derivada requiere la transformación reconocible de una obra preexistente y la identificación de un vínculo sustancial entre ambas. En el contexto de la IAG, esta exigencia

plantea serias dificultades prácticas, pues los modelos operan sobre millones de datos y generan resultados a partir de patrones estadísticos, lo que dificulta, mejor dicho, imposibilita, establecer qué obra concreta ha sido reproducida o transformada. No obstante, si pudiera demostrarse que un resultado constituye una reproducción sustancial o una transformación identificable de una obra determinada, sería necesario contar con la autorización previa del titular, conforme a lo dispuesto en el Convenio de Berna (1886) (art. 12) (Samuelson, 2023; United States Copyright Office, 2025).

Definida su naturaleza jurídica surge la cuestión de si los contenidos generados por la IAG deben recibir algún tipo de protección y, en su caso, bajo qué régimen. En el debate comparado pueden identificarse, al menos, tres enfoques principales. El primero propone reconocer protección únicamente a las contribuciones humanas identificables, especialmente en supuestos de creación híbrida en los que el usuario dirige de manera sustantiva el proceso creativo mediante decisiones relevantes. En estos casos, la protección se limita estrictamente a la aportación humana y no al resultado algorítmico en sí mismo. El segundo enfoque plantea la creación de un régimen *sui generis*, inspirado en el modelo europeo de protección de bases de datos, que otorgue derechos limitados de explotación a quienes realizan inversiones significativas en el desarrollo y entrenamiento de sistemas de IAG, sin equiparar dichos derechos a la autoría tradicional. Finalmente, una tercera postura sostiene que los contenidos generados de manera autónoma por la IAG deben permanecer en el dominio público, como parte de un acervo común que fomente la innovación, evite la sobreprotección y limite la concentración de poder en manos de grandes desarrolladores tecnológicos (Lucchi, 2025).

Cualquiera que sea la solución adoptada, resulta indispensable preservar un equilibrio entre la promoción de la innovación tecnológica y la protección de los intereses legítimos de los autores humanos cuyas obras nutren los *datasets* de entrenamiento. Una protección amplia e indiscriminada de los resultados algorítmicos podría desvirtuar la finalidad del derecho de autor y debilitar los incentivos para la creación humana, mientras que la ausencia absoluta de reconocimiento podría desalentar la inversión y el desarrollo tecnológico. En consecuencia, la literatura reciente coincide en la conveniencia de diseñar marcos normativos diferenciados: autoría exclusiva para obras humanas, protección limitada para creaciones híbridas y acceso abierto o protección *sui generis* para los resultados autónomos, siempre con pleno respeto a los derechos de los titulares de las obras preexistentes utilizadas en el entrenamiento de los sistemas de IAG (Comisión Europea, 2022; Samuelson, 2023; Lucchi, 2025).

Las distintas aproximaciones doctrinales y regulatorias en torno a la naturaleza jurídica y a la eventual protección del contenido generado por la IAG evidencian que no existe una solución única ni uniforme, sino una pluralidad de respuestas influenciadas por factores nacionales e internacionales. Esta diversidad de enfoques hace necesario adoptar una perspectiva comparada, que permita identificar cómo distintos ordenamientos han abordado, o comenzado a abordar, estos desafíos, así como evaluar, a la luz de dichas experiencias, las fortalezas, limitaciones y vacíos del marco jurídico mexicano frente al fenómeno tecnológico estudiado.

PERSPECTIVA COMPARADA Y ANÁLISIS DEL MARCO JURÍDICO MEXICANO

El tratamiento jurídico del uso de obras protegidas en *datasets* de entrenamiento y de los resultados generados por la IAG presenta diferencias significativas entre jurisdicciones, reflejo de enfoques regulatorios diversos frente a un mismo fenómeno tecnológico. Estas divergencias responden tanto a tradiciones jurídicas distintas como a prioridades de política pública relacionadas con la innovación, la protección de derechos fundamentales y la competitividad económica. Las jurisdicciones seleccionadas responden a modelos regulatorios representativos

y diferenciados frente al fenómeno analizado. Esta sección permite identificar convergencias, divergencias y posibles alternativas aplicables al contexto mexicano.

En Estados Unidos de América, el enfoque predominante mantiene el énfasis en la autoría humana como condición indispensable para la protección por derecho de autor. La *U.S. Copyright Office* (2023) ha reiterado de manera consistente que la protección se otorga únicamente a obras creadas por un ser humano, excluyendo del registro aquellos contenidos generados de forma autónoma por sistemas inteligentes. Esta posición quedó claramente reflejada en el caso *Zarya of the Dawn*, donde se determinó que solo los elementos textuales elaborados por la autora eran protegibles, mientras que las imágenes generadas mediante MidJourney carecían de protección por ausencia de intervención humana creativa suficiente. En lo relativo al uso de obras en *datasets* de entrenamiento, el sistema estadounidense no ha incorporado excepciones específicas y, en su lugar, la licitud de estos usos suele analizarse caso por caso bajo la doctrina del *fair use*. Esta flexibilidad ha permitido cierto margen para el desarrollo de tecnologías de la IAG, aunque a costa de una elevada incertidumbre jurídica para autores y desarrolladores (Samuelson, 2023; United States Copyright Office, 2025).

La Unión Europea, en contraste, ha adoptado un enfoque más preventivo y normativamente estructurado. La Directiva (UE) 2019/790 sobre derechos de autor en el mercado único digital introdujo excepciones obligatorias para la minería de textos y datos, permitiendo estas prácticas tanto con fines de investigación científica como, bajo ciertas condiciones, con fines comerciales, siempre que los titulares no hayan formulado una reserva expresa. Este régimen busca fomentar la innovación tecnológica sin vaciar de contenido los derechos exclusivos de los autores. A ello se suma el Reglamento de Inteligencia Artificial (*AI Act*), aprobado en 2024, que establece obligaciones transversales de transparencia, gobernanza y gestión de riesgos para los desarrolladores de sistemas de IA, incluyendo deberes de información sobre los datos utilizados en el entrenamiento. En conjunto, el modelo europeo refleja una estrategia orientada a equilibrar innovación, protección de derechos fundamentales y seguridad jurídica mediante una gobernanza proactiva de la tecnología (Comisión Europea, 2022; Lucchi, 2025).

En México, el panorama es considerablemente distinto, debido a que el marco jurídico vigente, encabezado por la Ley Federal del Derecho de Autor, no contempla expresamente la existencia de contenidos generados por IA. Por su parte, China ha optado por una estrategia intervencionista, caracterizada por un fuerte control estatal sobre el desarrollo y uso de la IA. A través de regulaciones emitidas desde 2022 y consolidadas en las *Interim Measures for the Management of Generative Artificial Intelligence Services* (2023), las autoridades chinas exigen a los proveedores de servicios generativos garantizar que los *datasets* de entrenamiento no infrinjan derechos de autor, registrar sus fuentes de datos y etiquetar de forma visible los contenidos generados por IA. Este enfoque responde a una concepción de la IA como recurso estratégico nacional, en la que la promoción de la innovación se encuentra estrechamente vinculada a objetivos de seguridad informativa, control social y soberanía tecnológica (Cyberspace Administration of China, 2023; Creemers, 2025).

El marco mexicano no establece excepciones específicas para la minería de textos y datos generados por IA. Aunque la ley reconoce de manera amplia los derechos morales y patrimoniales de los autores, no ofrece respuestas claras para escenarios de creación híbrida ni para el uso masivo de obras protegidas en procesos de entrenamiento algorítmico. Esta ausencia normativa genera un elevado grado de incertidumbre jurídica, ya que, en principio, la incorporación de obras protegidas en *datasets* podría calificarse como infracción, sin que existan mecanismos legales claros de excepción, compensación o licenciamiento colectivo. Asimismo, México carece de obligaciones específicas de transparencia o trazabilidad comparables a las previstas en el modelo europeo, lo que limita la capacidad de tutela efectiva de los derechos de autor en este contexto.

El análisis comparado permite identificar tendencias regulatorias diferenciadas en torno al entrenamiento de modelos generativos y la protección de los derechos de autor. Con el propósito de sistematizar las principales diferencias identificadas, en la tabla 1 se presenta un resumen de los elementos centrales de cada modelo jurídico analizado.

Tabla 1. Tendencias regulatorias comparadas

Jurisdicción	Uso de obras en <i>datasets</i>	Excepciones en minería de datos	Transparencia	Protección de resultados
Estados Unidos de América	Analizado bajo <i>fair use</i>	No de manera expresa	Limitada	Requiere intervención humana
Unión Europea	Permitido en ciertos supuestos	Sí, prevista en Directiva 2019/790	Obligaciones derivadas de la AI Act	Restrictiva respecto a autoría, a través de la IAG
China	Permitido bajo supervisión estatal	Regulación específica sobre IAG	Obligatoria y supervisada	Sujetos a control administrativo
México	Sin regulación específica	No prevista	No existen obligaciones específicas	Sin regulación específica

Fuente: elaboración propia con base en United States Copyright Office (2025), Comisión Europea (2022; 2025), Cyberspace Administration of China (2023) y Ley Federal del Derecho de Autor.

Como puede observarse en la tabla 1, las jurisdicciones analizadas han adoptado estrategias regulatorias diferenciadas frente al fenómeno estudiado. Frente a estas experiencias, México enfrenta el desafío de definir un marco normativo propio que le permita compatibilizar la protección de los autores con el fomento de la innovación tecnológica y el acceso al conocimiento. La experiencia comparada sugiere que avanzar hacia excepciones específicas para la minería de datos, incorporar obligaciones de transparencia en el uso de *datasets* y explorar esquemas de licenciamiento o remuneración colectiva, podrían constituir pasos clave para construir un régimen jurídico equilibrado y adaptado a los retos de la IAG.

ALTERNATIVAS REGULATORIAS PARA HACER FRENTE AL FENÓMENO – HALLAZGOS PRINCIPALES DEL ANÁLISIS COMPARADO

El análisis doctrinal comparado realizado permite identificar cuatro hallazgos principales. En primer lugar, existe consenso respecto a que el entrenamiento de los modelos generativos implica actos jurídicamente relevantes de reproducción y transformación de obras protegidas. En segundo término, las jurisdicciones analizadas han adoptado respuestas regulatorias diferenciadas: a) Estados Unidos privilegia un enfoque flexible basado en el *fair use*; b) la Unión Europea ha desarrollado un modelo preventivo sustentado en excepciones específicas y obligaciones de transparencia; c) China mantiene una estrategia de supervisión estatal más intensa; y d) México presenta importantes vacíos regulatorios.

El estudio confirma que la opacidad algorítmica constituye uno de los principales obstáculos para la tutela efectiva de los derechos de autor, debido a las dificultades para identificar las obras utilizadas en el entrenamiento de la IAG y establecer vínculos con los resultados generados. Finalmente, se observa una tendencia creciente hacia modelos regulatorios híbridos que combinan excepciones para minería de datos, obligaciones de transparencia y mecanismos de compensación o licenciamiento colectivo.

La experiencia comparada muestra que las respuestas más adecuadas frente al uso de obras protegidas en los *datasets* de entrenamiento de la IAG no se encuentran en la prohibición ni en la liberalización de dichas prácticas, sino en la adopción de modelos regulatorios híbridos que permitan equilibrar la protección de los derechos de autor con el fomento a la innovación tecnológica. En este sentido, la literatura reciente y los desarrollos normativos coinciden en la conveniencia de articular soluciones que combinen excepciones específicas, obligaciones de transparencia y esquemas de licenciamiento o remuneración colectiva, con el objetivo de reducir costos de transacción, fortalecer la tutela efectiva de los autores y proporcionar seguridad jurídica a los desarrolladores de tecnología (Lucchi, 2025; United States Copyright Office, 2025; Comisión Europea, 2022).

Una primera alternativa consiste en la incorporación de excepciones específicas para la minería de textos y datos, las cuales reconozcan que el entrenamiento de modelos de IAG requiere necesariamente operaciones técnicas de copia y análisis masivo de información. Estas excepciones deben estar cuidadosamente delimitadas, a fin de evitar afectaciones desproporcionadas a los derechos exclusivos de los autores (Ziaja, 2024; Lucchi, 2025).

Una segunda línea de acción prioritaria es el establecimiento de obligaciones de transparencia, trazabilidad y rendición de cuentas en relación con los *datasets* de entrenamiento y los procesos algorítmicos. La práctica ha evidenciado que, en ausencia de información mínima sobre la procedencia y composición de los datos utilizados, los derechos de autor resultan difícilmente exigibles. Por ello, los desarrollos regulatorios más recientes, particularmente en la Unión Europea, han avanzado hacia la imposición de deberes de documentación técnica, políticas explícitas de respeto a los derechos de autor y la elaboración de resúmenes públicos sobre el contenido utilizado en el entrenamiento de modelos de propósito general (Comisión Europea, 2025). Estas obligaciones buscan reducir la opacidad característica de la IAG y permitir que los titulares de derechos conozcan, al menos de manera razonable, si sus obras han sido utilizadas, bajo qué condiciones y con qué mecanismos de exclusión o reclamación cuentan.

Una tercera alternativa responde al problema práctico de los costos de transacción asociados al licenciamiento individual de millones de obras. En este contexto han cobrado relevancia los modelos de licenciamiento colectivo. Estos esquemas permiten autorizar el uso masivo de contenidos sin requerir autorizaciones individuales, a la vez que garantizan una compensación proporcional a los titulares mediante mecanismos institucionalizados de reparto. El Parlamento Europeo ha analizado expresamente la viabilidad de estas soluciones en el contexto de la IAG, destacando su potencial para compatibilizar la innovación y la retribución justa (Lucchi, 2025). En Estados Unidos, se ha reconocido que la discusión sobre el entrenamiento de modelos generativos se sitúa en la tensión entre la factibilidad de licenciar a gran escala y la preservación de los incentivos creativos, contemplando tanto las licencias colectivas como la aplicación flexible del *fair use* en determinados escenarios (United States Copyright Office, 2025).

A partir de estas tendencias, una ruta de implementación razonable para México podría ser la adopción de un enfoque escalonado y combinado. En primer lugar, sería pertinente introducir una excepción vinculada con la minería de datos, acotada y claramente definida, eventualmente acompañada de un mecanismo de reserva expresa para los titulares. En segundo término, resultaría indispensable incorporar obligaciones mínimas de transparencia y documentación sobre el uso de *datasets*, inspiradas en estándares comparados, con el fin de reducir las asimetrías informativas existentes. Por último, debería explorarse la habilitación de mecanismos de licenciamiento o remuneración colectiva, que permitan una retribución proporcional a los autores sin bloquear el desarrollo tecnológico, para lo cual podría echarse mano del entramado institucional y de organismos, que facilita el cobro de regalías a favor de

los autores. En conjunto, este enfoque mixto permitiría evitar dos escenarios indeseables: por un lado, que el derecho de autor se torne irrelevante por falta de mecanismos efectivos de cumplimiento; o, por el otro, que la innovación tecnológica se vea paralizada por exigencias de autorización individual imposibles de cumplir.

CONCLUSIONES

El análisis doctrinal, normativo y comparado desarrollado en esta investigación permitió identificar que el uso de obras protegidas por derecho de autor en los *datasets* de entrenamiento de la IAG constituye un fenómeno jurídicamente relevante, que desafía de manera directa los fundamentos tradicionales del régimen autoral. Los resultados obtenidos muestran que los procesos de entrenamiento implican, desde una perspectiva técnica y jurídica, actos de reproducción, fijación y, en determinados supuestos, de transformación de obras protegidas, los cuales no encuentran una respuesta clara ni homogénea en el marco normativo vigente a escala internacional. Lo anterior permite afirmar que el derecho de autor, tal como ha sido concebido, resulta insuficiente para atender la magnitud y complejidad de los usos masivos y automatizados propios de la IAG.

Asimismo, el análisis evidencia que la opacidad técnica de los modelos generativos constituye uno de los principales obstáculos para la tutela efectiva de los derechos de autor. La dificultad para rastrear las obras utilizadas en el entrenamiento y establecer una relación causal entre dichas obras y los resultados generados debilita los mecanismos tradicionales de protección, desplaza la carga probatoria hacia los titulares de derechos y profundiza las asimetrías existentes entre creadores y desarrolladores tecnológicos. Este hallazgo coincide con la literatura reciente, que ha advertido que la falta de trazabilidad y transparencia algorítmica erosiona la eficacia práctica del régimen autoral en entornos de la IA.

Desde una perspectiva comparada, los resultados permiten identificar respuestas regulatorias divergentes. Mientras que Estados Unidos ha optado por un enfoque flexible y casuístico apoyado en la doctrina del *fair use*, la Unión Europea ha avanzado hacia un modelo más estructurado mediante la introducción de excepciones específicas para la minería de textos y datos y obligaciones de transparencia asociadas al desarrollo de sistemas de la IA. China, por su parte, ha adoptado una estrategia intervencionista, basada en fuertes mecanismos de control estatal sobre los *datasets* y los contenidos generados. Frente a estos modelos, el análisis del marco jurídico mexicano pone de relieve un rezago normativo significativo, caracterizado por la ausencia de disposiciones específicas que regulen el uso de obras protegidas en los procesos de entrenamiento de la IAG, lo que genera incertidumbre jurídica tanto para autores como para desarrolladores.

En este contexto, el estudio concluye que para hacer frente a este fenómeno tecnológico no basta con la aplicación rígida de las categorías tradicionales del derecho de autor, sino que se requiere adoptar enfoques regulatorios híbridos. La experiencia comparada sugiere que la incorporación de excepciones acotadas para la minería de datos, el establecimiento de obligaciones de transparencia y trazabilidad en el uso de obras dentro de los *datasets* de entrenamiento, así como la exploración de esquemas de licenciamiento o remuneración colectiva, pueden contribuir a equilibrar la protección de los derechos de autor con el fomento de la innovación tecnológica. Estas alternativas permiten reducir costos de transacción, fortalecer la tutela de los creadores y proporcionar mayor seguridad jurídica, sin obstaculizar el desarrollo de la IAG.

Finalmente, este trabajo abre diversas líneas para futuras investigaciones. Resulta necesario profundizar en el diseño institucional de mecanismos de compensación colectiva aplicables al entrenamiento de modelos generativos, analizar la viabilidad técnica y jurídica de soluciones

de trazabilidad algorítmica, así como estudiar el impacto de estos modelos regulatorios desde una perspectiva de derechos humanos. Asimismo, futuras investigaciones podrían explorar empíricamente el impacto económico y cultural de la IAG sobre las industrias creativas, con el fin de aportar evidencia adicional que oriente la formulación de políticas públicas y reformas legislativas en esta materia.

REFERENCIAS

- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, V., Card, D., Castellon, R., Chatterji, N. S., Chen, A., Creel, K. A., Davis, J. Q., Demszky, D., ... Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258. <https://arxiv.org/abs/2108.07258>
- Buick, A. (2024). Copyright and AI training data—Transparency to the rescue? *Journal of Intellectual Property Law & Practice*, 20(3), 182-192. <https://doi.org/10.1093/jiplp/jpae102>
- Comisión Europea. (2022). *Study on copyright and artificial intelligence*. Publications Office of the European Union.
- (2025). *General-purpose AI Code of Practice*. Publications Office of the European Union. <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai>
- Convenio de Berna para la Protección de las Obras Literarias y Artísticas (1886). <https://www.wipo.int/treaties/es/ip/berne/>
- Creemers, R. (2025). *The regulation of generative AI in China*. Social Science Research Network. <https://ssrn.com/abstract=5228697>
- Cyberspace Administration of China (CAC). (2023). *Interim Measures for the Management of Generative Artificial Intelligence Services*. Cyberspace Administration of China.
- De la Parra, E. (2015). *Derechos humanos y derechos de autor. Las restricciones al derecho de explotación* (Universidad Nacional Autónoma de México e Instituto de Investigaciones Jurídicas, Eds.; 2ª ed.).
- Directiva 2001/29/CE del Parlamento Europeo y del Consejo, de 22 de mayo de 2001, sobre la armonización de determinados aspectos de los derechos de autor y derechos afines en la sociedad de la información. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32001L0029>
- Directiva (UE) 2019/790 del Parlamento Europeo y del Consejo, 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. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32019L0790>
- Gervais, D. J. (2020). The machine as author. *Iowa Law Review*, 105(5), 2053-2106.
- Ginsburg, J. C., & Budiardjo, L. (2019). Authors and machines. *Berkeley Technology Law Journal*, 34, 343-448.
- Henderson, P., Li, X., Jurafsky, D., Hashimoto, T., Lemley, M. A., & Liang, P. (2023). Foundation models and fair use. *Journal of Machine Learning Research*, 24, 400, 1-79. <https://arxiv.org/abs/2303.15715>
- Kretschmer, M., Margoni, T., & Oruç, P. (2024). Copyright law and the lifecycle of machine learning models. *IIC — International Review of Intellectual Property and Competition Law*, 55(1), 110-138. <https://doi.org/10.1007/s40319-023-01419-3>
- Lipszyc, D. (1993). *Derecho de autor y derechos conexos* (Centro Regional para el Fomento del Libro en América Latina y el Caribe [Cerlalc], Ed.).

- Lucchi, N. (2025). *Generative AI and copyright: Training, creation and regulation* (Study No. PE 774.095). Policy Department for Justice, Civil Liberties and Institutional Affairs, European Parliament. [https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU\(2025\)774095_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)774095_EN.pdf)
- Mitchell, T. (2025). Trust and transparency in artificial intelligence. *Philosophy & Technology*, 38(3), 1-30. <https://doi.org/10.1007/s13347-025-00916-2>
- Pasetti, M., Santos, J. W., Corrêa, N. K., De Oliveira, N., & Barbosa, C. P. (2025). Technical, legal, and ethical challenges of generative artificial intelligence: An analysis of the governance of training data and copyrights. *Discover Artificial Intelligence*, 5, Article 193. <https://doi.org/10.1007/s44163-025-00379-6>
- Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Acta Legislativa: Artificial Intelligence Act). *Diario Oficial de la Unión Europea*, L 1689, 12.7.2024. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th Ed.). Pearson.
- Samuelson, P. (2023). Generative AI meets copyright. *Science*, 381(6654), 158-161. <https://doi.org/10.1126/science.adi0656>
- Sengar, S. S., Hasan, A. B., Kumar, S., & Carroll, F. (2024). Generative artificial intelligence: A systematic review and applications. *Multimedia Tools and Applications*, 84(21), 23661-23700. <https://doi.org/10.1007/s11042-024-20016-1>
- Tribunal de Justicia de la Unión Europea. (2009). Infopaq International A/S v. Danske Dagblades Forening (Asunto C-5/08). ECLI:EU:C:2009:465. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:62008CJ0005>
- United States Copyright Office. (2023). *Copyright registration guidance: Works containing AI-generated material*. U.S. Copyright Office.
- (2025). *Copyright and artificial intelligence, Part 3: Generative AI training (pre-publication version)*. U.S. Copyright Office. <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>
- Ziaja, G. M. (2024) The text and data mining opt-out in Article 4(3) CDSMD: Adequate veto right for rightholders or a suffocating blanket for European artificial intelligence innovations? *Journal of Intellectual Property Law and Practice*, 19(5), 453-459.

Financiación

El presente trabajo no cuenta con financiación específica de agencias de financiamiento en los sectores público o privado para su desarrollo y/o publicación.

Conflicto de interés

El autor del trabajo declara no tener ningún conflicto de intereses en su realización.

Contribución de autoría

El autor fue responsable de la conceptualización, diseño metodológico, recopilación y análisis de información, redacción y revisión final del manuscrito.

