

# DATASETS DE ENTRENAMIENTO EN LA INTELIGENCIA ARTIFICIAL GENERATIVA: DESAFÍOS PARA EL DERECHO DE AUTOR

## AUTOR

César Ricardo Castillo Velazco<sup>1</sup>

<sup>1</sup> Centro de Ingeniería y Desarrollo Industrial (CIDESI).

✉ cesar.castillo@cidesi.edu.mx

ID <https://orcid.org/0000-0002-1130-1816>

Artículo de investigación

Revista Especializada en Investigación Jurídica  
Vol. 10, Núm. 19  
julio-diciembre 2026

ISSN: 2448-8739 (En línea)

DOI: <https://doi.org/10.20983/rej.2026.2.1>

*Training datasets in generative artificial intelligence: challenges for copyright*

📅 Recibido: 25/DICIEMBRE/2025

✅ Aprobado: 22/MAYO/2026

🌐 Publicación online: 02/JUNIO/2026

## ¿DE QUÉ TRATA ESTA INVESTIGACIÓN?



Analiza las implicaciones jurídicas del uso de obras protegidas por derecho de autor en los datasets de entrenamiento de la inteligencia artificial generativa (IAG), con el fin de identificar las principales tensiones normativas y brechas regulatorias existentes.

## METODOLOGÍA



Revisión doctrinal

+



Revisión normativa

+



Perspectiva de derecho comparado

Estudio cualitativo que permite analizar y comparar el tratamiento jurídico en distintas jurisdicciones.

## HALLAZGOS CLAVE



### 1. USO MASIVO DE OBRAS PROTEGIDAS

El entrenamiento de modelos generativos implica actos jurídicamente relevantes de reproducción y transformación que no encuentran una respuesta clara en el marco jurídico internacional vigente.



### 2. OPACIDAD TÉCNICA

La opacidad algorítmica dificulta la trazabilidad de las obras utilizadas y limita el ejercicio efectivo de los derechos de autor.



### 3. INCERTIDUMBRE JURÍDICA

La incorporación de obras protegidas en datasets podría calificarse como infracción, sin que existan mecanismos claros de excepción, compensación o licenciamiento colectivo.



### 4. BRECHAS REGULATORIAS

Existen diferencias significativas entre jurisdicciones respecto al uso de obras, excepciones para minería de datos, transparencia y protección de resultados.



## CONCLUSIÓN

Es necesario 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 y fomentando la **innovación responsable**.



## PALABRAS CLAVE

- datasets de entrenamiento
- derecho de autor
- inteligencia artificial generativa

## TENDENCIAS REGULATORIAS COMPARADAS

JURISDICCIÓN	USO DE OBRAS EN DATASETS	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.

## ¿POR QUÉ ES IMPORTANTE?



El debate actual exige adaptar el derecho de autor a los avances tecnológicos.



Debe equilibrarse la protección de los autores con el impulso a la innovación.



La transparencia y la rendición de cuentas son clave para garantizar derechos y confianza.



Un marco jurídico adecuado puede fomentar la innovación responsable y sostenible.