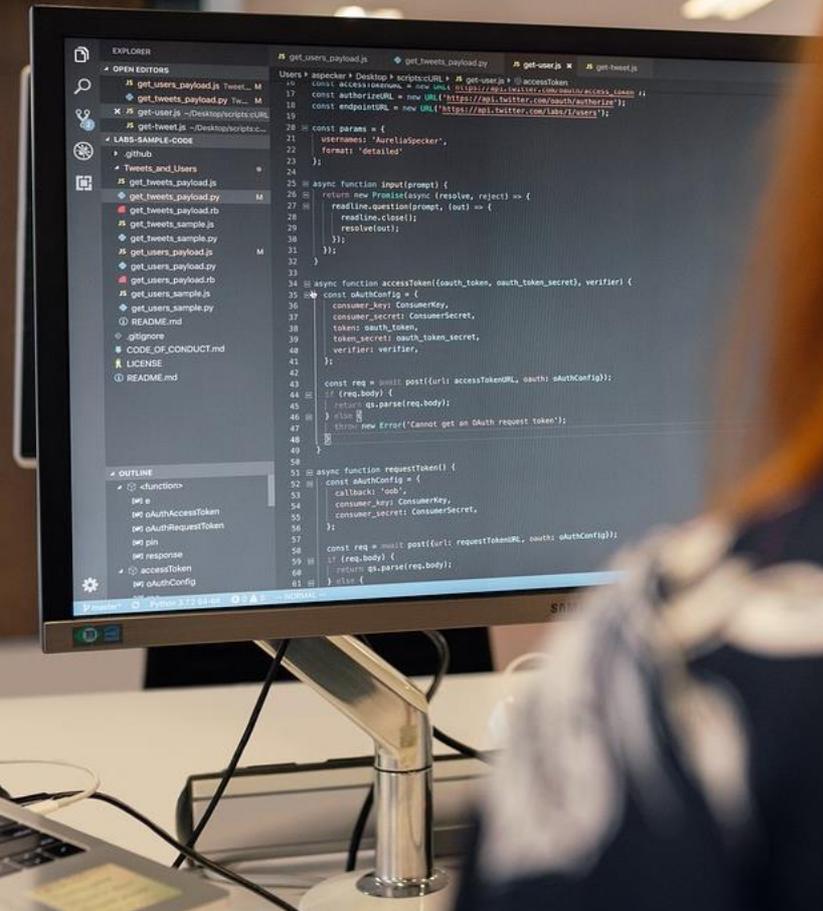


NOTA INFORMATIVA

Inteligencia artificial,
minería de datos y
derechos de autor



Primera sentencia en Europa sobre este complejo equilibrio.



Lo que necesitas saber

- **Definición del alcance de la excepción de interés científico en el ámbito del *text and data mining*:** en base a esta sentencia, los derechos de autor quedan limitados por cuestiones de interés público. En concreto, se refuerza la ya establecida excepción por uso científico y de investigación. Así las cosas, los *datasets*, pueden incluir obras protegidas sin necesidad de licencias individuales para el entrenamiento de IA generativa, con ciertas limitaciones.
- **Impacto en la industria creativa:** la sentencia podría establecer un precedente que limite las posibilidades de que los autores controlen cómo sus obras pueden ser transformadas en otras obras derivadas mediante la IA generativa.
- **Aplicación de la Directiva 2019/790 de Derechos de Autor en el Mercado Único Digital (DSM) en el ámbito de la IA generativa:** a pesar de que el fallo se base en derecho alemán, el fundamento de esta regulación reside en los artículos 4 y 5 de la DSM, en relación con la minería de datos y las excepciones aplicables a sus restricciones correspondientes.





El pasado 27 de septiembre de 2024, el Tribunal Regional de Hamburgo dictó la primera sentencia en Europa que aborda uno de los temas más controvertidos en el contexto de *text and data mining (TDM)*: la confrontación entre IA y derechos de autor. Con esta sentencia, el Tribunal analiza la legalidad del uso de imágenes protegidas por derechos de autor en conjuntos de datos (los llamados *datasets*) para entrenar modelos de IA generativa.

El litigio se inicia por una demanda del fotógrafo alemán **Robert Kneschke**, quien alegó que una organización (**LION**) había utilizado su fotografía sin su autorización para crear un *dataset* público cuya finalidad es entrenar sistemas de IA generativa.

Por su parte, la entidad demandada LION, es una organización sin ánimo de lucro compuesta por investigadores, comprometida con el desarrollo de **algoritmos de autoaprendizaje** en el contexto de la IA. Kneschke alegó que LION había incluido su fotografía en un conjunto de datos masivo (de 5,85 mil millones de pares imagen-texto) extraído de diversas fuentes de internet. La fotografía en cuestión estaba integrada por voluntad de Kneschke en una plataforma de imágenes, en cuyos términos se establecía que la imagen estaba protegida por derechos de autor y se distribuía bajo licencia, con **restricciones explícitas** en contra del *scraping* de datos mediante programas automatizados.

Concretamente, en esta plataforma, constaba una cláusula que establecía expresamente: “No se puede utilizar programas automatizados, *applets*, *bots* o similares para acceder a este sitio web o cualquier contenido en él para cualquier propósito, incluyendo, a modo de ejemplo, descargar contenido, indexar, extraer o almacenar cualquier contenido del sitio web”.

LION, obviando esta restricción, **descargó** de la plataforma, sin autorización, la fotografía, la incluyó en su *dataset* y la utilizó en un **proceso automatizado de análisis**, comparando su contenido con descripciones textuales para validar su precisión. Estas acciones implican la descarga, **reproducción** y, según alega Kneschke, el uso no autorizado de la fotografía. Concretamente, Kneschke argumentó que la inclusión de su fotografía en el conjunto de datos de Lion infringe específicamente su derecho exclusivo de reproducción, ya que ésta fue descargada y almacenada por LION sin su consentimiento para integrarse en su *dataset*.

¿Están cubiertas estas acciones por las excepciones legales relacionadas con la minería de datos contempladas por la DSM?

(I) El fallo del Tribunal

El Tribunal reconoció que LION había realizado una reproducción de la fotografía protegida, pero determinó que dicha acción estaba amparada por la excepción legal prevista en el **Artículo 60d de la UrhG** (Ley de Derechos de Autor alemana), que permite el uso de obras protegidas, en el ámbito de la minería de datos para fines de investigación científica.

Esta excepción, traspone el Artículo 4 de la DSM, permitiendo la minería de datos siempre que no exista una reserva válida de derechos que lo prohíba.

El Tribunal interpretó que la fotografía extraída por LION cumplía con los requisitos de accesibilidad legal, al estar destinada a la investigación, sin perseguir ningún tipo de fin comercial.

Por ello, el Tribunal aceptó el argumento de LION consistente en que actuaba dentro de los márgenes de la **investigación científica no comercial**, lo que le permitía utilizar



temporalmente la fotografía sin necesidad de obtener una autorización expresa del fotógrafo. Cabe destacar que LION no genera outputs con las imágenes de su *dataset*, sino que simplemente las utiliza en sus análisis para obtener correlaciones entre imagen y texto.

En cuanto a la cláusula de "reserva de derechos" (el conocido *opt-out*), el Tribunal abordó si Kneschke podía acogerse a la restricción que prohibía expresamente, en la plataforma en la que estaba incorporada la fotografía, el uso automatizado de sus contenidos. Aquí entra en juego el debate sobre si esta reserva cumple con los requisitos técnicos de legibilidad por máquina, como lo exige la **DSM**, que estipula que cualquier *opt-out* debe ser formulado **de manera adecuada, como medios de lectura mecánica**, o, en otras palabras: de forma que sea comprensible para los sistemas automatizados.

Sin embargo, el Tribunal no innovó en este aspecto y no ratificó que la reserva alegada cumpliera con los estándares técnicos necesarios. Si bien el Tribunal hizo un amago de aceptar que la reserva utilizada en las condiciones de uso del sitio web en lenguaje natural podría encajar en el requisito legal de una reserva en formato legible por máquina, finalmente se limitó a aceptar que se deberá valorar esta cuestión *ad casum*. De hecho, señaló que parece haber cierta contradicción al permitir a los proveedores de modelos de IA desarrollar sistemas de IA de comprensión y creación de textos cada vez más potentes y sofisticados, por un lado, pero no exigirles que utilicen estos modelos para detectar las reservas de los titulares de derechos de *opt-out* hechas con lenguaje natural, por otro. Sin embargo, dado que la sentencia finalmente se resuelve en base a la excepción por interés científico y no en función de la validez del *opt-out*, esta cuestión resulta, en última instancia, irrelevante para el fallo en sí.

Además, el Tribunal se refirió al **Reglamento de Inteligencia Artificial (Reglamento (UE) 2024/1689)**, que introduce nuevos requisitos para el uso de *datasets* en el entrenamiento de IA. Según el artículo 53.1(c), los desarrolladores de IA deben tener estrategias claras para identificar y respetar cualquier reserva de derechos válida, como las previstas en el artículo 4 de la DSM.

No obstante, aunque el Reglamento de IA es relevante para el contexto normativo general, en este caso no fue decisivo para el fallo, que se centró en la interpretación de la **Directiva DSM** y las excepciones para investigación científica establecidas en la ley alemana.

(II) Conclusión

El fallo del Tribunal Regional de Hamburgo es pionero en la regulación del uso de obras protegidas en *datasets* para entrenar IA generativa en Europa. No obstante, **ni ha supuesto un cambio de paradigma ni debe ser asumida como jurisprudencia**. No debemos olvidar que es muy probable que esta sentencia trascienda al Tribunal de Apelación de Hamburgo, al Tribunal Federal de Justicia alemán e incluso al Tribunal de Justicia Europeo.

Asimismo, si bien beneficia a organizaciones de investigación científica, plantea desafíos significativos para los titulares de derechos de autor que buscan proteger sus obras en un entorno donde la tecnología avanza a pasos agigantados.

La interpretación de la excepción de interés científico que ha hecho el Tribunal de Hamburgo es, innegablemente, flexible, ya que se centró esencialmente en la ausencia de actividad comercial y en que el conjunto de datos se puso a libre disposición del público. El hecho de que empresas con actividad comercial también puedan utilizar el *dataset* creado por LION fue irrelevante para el fallo.



Este enfoque es, asimismo, ciertamente ambiguo y pone de manifiesto la necesidad de alcanzar en Europa una **definición unificada** de la excepción de interés científico. De hecho, a pesar de que nuestra Ley de Propiedad Intelectual sí incide en mayor medida en qué se consideran fines educativos o de investigación científica, en el contexto de la minería de datos nuestra regulación también resulta insuficiente. Las excepciones a la DSM no deben convertirse en una herramienta de blanqueo para opacar usos ilegítimos de derechos de propiedad intelectual. En base a la sentencia analizada, aquellas obras protegidas que entran en un *dataset* bajo el pretexto de un fin científico pueden perder la trazabilidad sobre cómo se integran en el mismo y qué outputs se generan a través de estas. De hecho, estos outputs sí que podrían escaparse de la excepción de interés científico, perseguir un fin comercial e incluso ser **competidores** de la obra originaria y protegida.

A todo ello, se suma el hecho de que tampoco estén claros los requisitos para hacer un *opt-out* efectivo, si bien no puede culparse al Tribunal de Hamburgo por ello. De hecho, actualmente, la Comisión Europea se encuentra elaborando un **Código de Buenas Prácticas** con los principales proveedores de Modelos de IA de Uso General y de IA generativa, con el fin de alcanzar un criterio unificado sobre los mecanismos técnicos de *opt-out*, iniciativa que habrá que seguir de cerca para analizar cómo se aplicarán estos requisitos en nuestro ordenamiento.

En cualquier caso, el debate sobre los derechos de autor en la era de la IA está lejos de concluir. Veremos si este caso trasciende y la jurisprudencia del TJUE aclara cómo pueden conciliarse los derechos de los creadores con el avance tecnológico que representa la IA. La cual, indudablemente, requiere de contenidos actualizados y de calidad para producir resultados de calidad.

Área de Tecnología, Medios y Telecomunicaciones de ECIJA

info@ecija.com

Telf: + 34 91.781.61.60