

El rol de la excepción de minerías de datos en la inteligencia artificial (IA) generativa – perspectiva europea¹

The role of the data mining exception in generative artificial intelligence (AI) – an european perspective

DR AVV. MAGALI CONTARDI²

RESUMEN

Los avances en las tecnologías de inteligencia artificial (IA) han impulsado el desarrollo de la IA generativa, capaz de producir una amplia gama de contenidos. No obstante, el entrenamiento de estos sistemas plantea importantes interrogantes sobre la salvaguardia y el ejercicio de los derechos de autor de los contenidos utilizados para estos fines. En este contexto, este artículo se propone analizar el rol y la idoneidad de la excepción de minería de datos ('Text and Data Mining') en el ámbito de la IA generativa, centrándose en el marco legislativo de la Unión Europea (UE). El artículo concluye argumentando que, aunque la minería de datos posee un potencial considerable en el desarrollo de la IA y su incorporación en el marco legislativo de la UE es crucial, tanto la estructura misma de la excepción como la diversidad de

1 Fecha de recepción: 16 de junio de 2024. Fecha de aceptación: 20 de noviembre de 2024. Received: June 16, 2024. Accepted: November 20, 2024. Artículo de revisión. DOI: <https://doi.org/10.18601/16923960.v24n1.02>

2 Doctora en Derecho. Universidad de Alicante, España. Abogada ('avvocato') especializada en derecho de la propiedad intelectual e industrial y de las nuevas tecnologías, miembro del colegio de abogados de Milán, Italia y Múnich, Alemania, fundadora de Contardi IP Law. Investigadora postdoctoral, Scuola Superiore Sant'Anna Pisa, Italia; investigadora, GIP-Law, Universidad de Alicante, España. Contacto: avv.mcontardi@gmail.com. Orcid Id: <https://orcid.org/0000-0002-3346-1570>

enfoques nacionales no ofrecen un cuadro normativo que permita una aplicación armonizada en la UE y ello podría socavar su uso en el campo de la IA generativa.

Palabras clave: Excepciones; minería de datos; inteligencia artificial; derechos de autor; Unión Europea; mercado único digital.

ABSTRACT

Advances in artificial intelligence (AI) technologies have driven the development of generative AI, capable of producing a wide range of content. However, the training of these systems raises significant questions regarding the protection and enforcement of copyright for the content used for these purposes. In this context, this article aims to analyze the role and suitability of the data mining exception ('Text and Data Mining') in the field of generative AI, focusing on the legislative framework of the European Union (EU). The article concludes by arguing that although data mining has considerable potential for AI development and its integration into the EU legislative framework is crucial, both the structure of the exception and the diversity of national approaches do not provide a regulatory framework that allows for its harmonized application across the EU, and this could undermine its use in the GenAI field.

Keywords: Exceptions, data mining, artificial intelligence, copyright, european union, digital single market

1. INTRODUCCIÓN: DERECHOS DE AUTOR, MINERÍA DE DATOS E INTELIGENCIA ARTIFICIAL, UN CONNUBIO COMPLEJO

En los últimos años, los avances significativos en las tecnologías en la Inteligencia Artificial (IA) han impulsado a la IA generativa, donde dichos sistemas pueden producir una gran variedad de contenido, desde la generación de artículos, fotografías y composiciones musicales hasta la creación de obras de arte. Si bien es en las industrias creativas donde la IA generativa ha hecho su mayor irrupción, esta puede además utilizarse en áreas como la salud, publicidad y marketing, manufactura, y la industria automotriz.³

3 Véase, indicativamente: World Intellectual Property Organization, 'Generative AI: Navigating Intellectual Property' (Geneva, Switzerland: World Intellectual Property

En términos simples, los modelos de IA generativa dependen de una gran cantidad de contenido, que utilizan para entrenarlos. Un aspecto fundamental en el funcionamiento de la IA Generativa es el proceso de lectura automática de grandes volúmenes de datos para descubrir patrones, generar nuevo conocimiento o extraer ideas. Estos sistemas se entrenan y desarrollan utilizando algoritmos de aprendizaje automático que analizan grandes cantidades de datos, imágenes y contenido ('datos de entrada o input'), y aprenden a utilizar esa información para crear nuevo contenido similar al contenido existente ('dato de salida u output').⁴ Se consideren, por ejemplo, los datos utilizados para entrenar la IA para los vehículos, que pueden incluir millones de fotografías de señales de tránsito, protegidas por derechos de autor o derechos conexos como fotografías.⁵ De modo análogo, la IA de aplicaciones musicales se basan en grabaciones de audio protegidas por los derechos de autor de los compositores y los derechos conexos de los artistas intérpretes y los productores discográficos.⁶ Dicho análisis, conocido también como técnicas de extracción de datos o minería de datos,⁷ podría considerarse como una reproducción, aunque sea temporal, de los datos

Organization (WIPO), 28 February 2024); World Intellectual Property Organization, 'The Frontier of MedTech: AI-Driven Medical Devices', WIPO, 20 November 2023, https://www.wipo.int/policy/en/news/global_health/2023/news_0021.html; Jan Hejtmanek, 'Early Generative AI and Its Impact on Automotive Industry, 2023 Summary', Deloitte Czech Republic, 2023, <https://www2.deloitte.com/cz/en/pages/technology/articles/early-generative-ai-and-its-impact-on-automotive-industry.html>.

- 4 Para un detalle sobre el funcionamiento de los algoritmos de IA generativa, véase: Andrés Guadamuz, 'A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs', SSRN Scholarly Paper (Rochester, NY, 26 February 2023), <https://doi.org/10.2139/ssrn.4371204>. Véase también: Nicola Lucchi, 'ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems', *European Journal of Risk Regulation*, 29 August 2023, 1–23, <https://doi.org/10.1017/err.2023.59>. (Detallando el funcionamiento de IA generativa de ChatGPT).
- 5 Paul Christianson, 'Billions of Miles of Data: The Autonomous Vehicle Training Conundrum', 30 September 2020, consultado el 10 Octubre 2024, <https://blog.cloudfactory.com/autonomous-vehicle-training-conundrum>.
- 6 Morgan Stevens, 'Training Musical AI Systems', *Center for Data Innovation* (blog), 30 June 2023, consultado el 10 Octubre 2024, <https://datainnovation.org/2023/06/training-musical-ai-systems/>.
- 7 Commission Staff Working Document Executive Summary of the Impact Assessment on the modernisation of EU copyright rules Accompanying the document Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market and Proposal for a Regulation of the European Parliament and of the Council laying down rules on the exercise of copyright and related rights applicable to certain online transmissions of broadcasting organisations and retransmissions of television and radio programmes SWD/2016/0302 final–2016/0284 (COD) §4.3.1. (Definiendo la minería de datos como del siguiente modo: "Text and Data Mining [TDM] is a term commonly used to describe the automated processing ("machine reading") of large volumes of text and data to uncover new knowledge or insights").

y contenidos utilizados, incluyendo cualquier obra protegida o porciones completas de las bases de datos empleadas.

Es por este motivo que el uso de fotos, imágenes o textos por los modelos de IA mediante técnicas de extracción de datos para el entrenamiento de la IA plantean interrogantes sobre la protección y ejercicio de los derechos de autor respecto de los contenidos que han sido utilizados con tal fin. Además, el uso de dicho contenido también podría generar conflicto con el derecho del creador de una base de datos para prohibir la extracción o reutilización de la totalidad o una parte sustancial de la misma.⁸ La cuestión fundamental a dirimir gira en torno a si el uso no autorizado de dichas obras protegidas por derechos de autor para el entrenamiento o aprendizaje de los modelos de IA constituye, o no, un acto de infracción a los derechos de autor. Cabe también plantearse si los resultados generados por la IA a partir de dichos 'input' infringen los derechos de los titulares del contenido utilizado como dato de entrada (o 'input'). En este sentido, se argumenta que, en la medida que el resultado final ('output') refleja el 'input', dicho 'output' podría constituir, a la vez, un ulterior acto de reproducción, como así también de puesta a disposición y, en algunos casos, de distribución.⁹ En los Estados Unidos (EE.

- 8 European Parliament, 'REPORT on intellectual property rights for the development of artificial intelligence technologies' (2 October 2020). Disponible: < https://www.europarl.europa.eu/doceo/document/A-9-2020-0176_EN.html>. En doctrina, se registra un creciente interés sobre el asunto. Véase, indicativamente: Mark A. Lemley, 'How Generative AI Turns Copyright Upside Down', SSRN Scholarly Paper (Rochester, NY, 21 July 2023), <https://doi.org/10.2139/ssrn.4517702>; Enrico Bonadio, Plamen Dinev, and Luke McDonagh, 'Can Artificial Intelligence Infringe Copyright? Some Reflections', in *Research Handbook on Intellectual Property and Artificial Intelligence*, ed. Ryan Abbott, Law (Cheltenham, UK: Edward Elgar Publishing, 2022), 245–57, <https://doi.org/10.4337/9781800881907.00007>; Guadamuz, 'A Scanner Darkly'; Samuelson Pamela, 'Legal Challenges to Generative AI, Part I', *Communications of the ACM*, 22 June 2023, <https://doi.org/10.1145/3597151>; Jenny Quang, 'Does Training AI Violate Copyright Law?', *Berkeley Technology Law Journal* 36 (2021): 1407; Lucchi, 'ChatGPT'.
- 9 Dos pronunciamientos del Tribunal de Justicia de la Unión Europea (TJUE) podrían tener una relevancia particular. La primera es la sentencia relativa al caso Infopaq (Sentencia del Tribunal de Justicia del 16 julio 2009, Infopaq International A/S y Danske Dagblades Forening, C-5/08, ECLI:EU:C:2009:465), mencionada anteriormente, en la que el TJUE afirmó que una secuencia de 11 palabras puede estar protegida por derechos de autor, siempre que dicha secuencia sea original (y, por lo tanto, su uso requiere la autorización del autor). La segunda sentencia relevante es la del caso Pelham (Sentencia del Tribunal de Justicia del 29 julio 2019, Pelham GmbH y Ralf Hütter y Florian Schneider-Esleben, C476/17, ECLI:EU:C:2019:624), en la que el Tribunal de Justicia de la Unión Europea (TJUE) estableció que el uso de una breve secuencia rítmica, conocida como "sampling", puede constituir una violación de los derechos exclusivos de los productores de fonogramas, cuando la muestra sonora sea reconocible al oído. Véase: Contardi, Magali, *Intelligenza Artificiale Generativa: creatività o violazione di diritti?* – Prospettiva Europea, *AIPPI News Italia*, julio 2023, consultado el 10 Octubre 2024, <https://aippi.it/newsletter/aippi-news-luglio-2023>.

UU) ya proliferan las demandas de infracción contra empresas generadoras de contenido artístico-visual a través de la IA.¹⁰ En la Unión Europea, en cambio, solo se registra un litigio judicial al respecto, cuya sentencia deja interrogantes sin resolver.¹¹

En este contexto, el presente artículo se propone de analizar cuál es el rol – obstáculo o facilitador – de la excepción de minería de datos (o, su acrónimo en inglés, Text and Data Mining, 'TDM') en el ámbito de la IA generativa, con un enfoque en la legislación de la Unión Europea (UE). Como es sabido, en el 2019, la Unión Europea adoptó la Directiva sobre Derechos de Autor y Derechos Conexos en el Mercado Único Digital (la Directiva DDAMUD),¹² la cual introdujo dos nuevas excepciones de derechos de autor para la minería de datos. La minería de texto y datos puede ser una herramienta útil en campos tan diversos como la investigación científica, el periodismo, la cultura y, no menos importante, el entrenamiento de la IA. Dada la gran cantidad de datos que utilizan los sistemas de IA para generar

- 10 Véase: Getty Images (US), Inc. v. Stability AI, Inc., No. 1:23-cv-00135-GBW (D. Del. Mar. 29, 2023); Silverman et al. v. OpenAI, Inc. et al., No. 4:23-cv-03416 (N.D. Cal. Jul. 7, 2023); Tremblay et al. v. OpenAI, Inc. et al., No. 4:2023-cv-03223 (N.D. Cal. Jul. 7, 2023). Estas demandas alegan que la incorporación de datos de entrenamiento por parte de modelos de IA generativa constituye una infracción de los derechos de autor de los titulares cuyos contenidos se han utilizado. Véase también: Magali Contardi, 'IA bajo acusa! ¿el uso de imágenes en el aprendizaje de la inteligencia artificial viola los derechos de autor?: implicancias para la creatividad | LVCENTINVS', IP Blog, *Lvcentinvs* (blog), 9 Marzo 2023, consultado el 10 Octubre 2024, <https://www.lvcentinvs.es/2023/03/09/ia-bajo-acusa-el-uso-de-imagenes-en-el-aprendizaje-de-la-inteligencia-artificial-viola-los-derechos-de-autor-implicancias-para-la-creatividad/>.
- 11 Véase, Sentencia del Tribunal de Hamburgo (Landgericht Hamburg), Kneschke, R. / LAION gemeinnütziger e. V, 310 O 227/23 del 27.09.2024. En Alemania, un fotógrafo de stock demandó a LAION, una organización sin fines de lucro con sede en Alemania que compila conjuntos de datos de entrenamiento para empresas tecnológicas que desean entrenar IA generativa. Stability AI utilizó el conjunto de datos LAION 5B para entrenar Stable Diffusion. A fines de septiembre, el tribunal de Hamburgo dictaminó que LAION podía acogerse a la excepción contemplada en la Sección 60d de la UrhG (norma que incorpora la excepción de minería de datos y texto para fines de investigación científica en la legislación Alemana). Sin embargo, esta decisión ha recibido críticas por parte de la comunidad académica quienes han planteado dudas sobre la idoneidad de este precedente jurisprudencial para servir de base a la IA generativa. Entre otros aspectos, se ha criticado la interpretación realizada por el tribunal de conceptos clave de la disposición, como 'organizaciones de investigación' e 'investigación científica'. Véase: Eleonora Rosati, 'The German LAION Decision: A Problematic Understanding of the Scope of the TDM Copyright Exceptions and the Transition from TDM to AI Training', *The IPKat* (blog), consultado el 15 Octubre 2024, <https://ipkitten.blogspot.com/2024/10/the-german-laion-decision-problematic.html>.
- 12 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 [2019], (Directiva (UE) 2019/790).

nuevo contenido, es evidente la estrecha relación entre la IA generativa y la excepción de minería de datos: la excepción de minería de texto y datos permite a los sistemas de IA acceder a grandes cantidades de datos, que son utilizados por la IA generativa para crear nuevo contenido. Así que, sin acceso a grandes volúmenes de datos, los algoritmos de IA tendrían dificultades para "aprender" y mejorar su rendimiento. Por lo tanto, si a estos sistemas no se les permitiera acceder a tales datos, su capacidad para generar contenido sin duda se vería limitada.¹³

Además, la Unión Europea ha recientemente adoptado un reglamento para regular la IA a nivel europeo, el cual incluye algunos artículos que podrían afectar la amplitud y aplicación de la excepción de minería de datos.¹⁴ Dados estos importantes cambios legislativos en la regulación de la IA relacionadas con la minería de datos y la prometedora utilidad de esta técnica es pertinente indagar como estas intervenciones legislativas se interrelacionan.

El resto del artículo se estructura de la siguiente manera: La siguiente parte (sección 2) ilustrará la excepción de minería de datos delineada en el texto de la directiva DDAMUD (sección 2.1), analizando, con un método de análisis normativo y teleológico, complementados con el análisis de literatura, si la formulación adoptada por el legislador europeo es idónea para alcanzar los objetivos perseguidos. Se presentará, además, un análisis exhaustivo de las distintas formulaciones adoptadas por los estados miembros de la Unión Europea en sede de incorporación de la directiva en su derecho interno, empleando un análisis comparativo, a los fines de complementar el análisis sobre la robusteza, rol e idoneidad de la excepción de minería de datos en el *aquis comunitario* en el contexto de la IA generativa. La siguiente parte (sección 3) volcará su atención al reglamento que regula los usos de la IA en la Unión Europea, incorporando, a la vez, algunos ejemplos de reglamentación de la

13 En este sentido, algunos afirman que "el futuro de la IA depende de la minería de datos y su capacidad para extraer y analizar datos a gran escala". Véase: P. Bernt Hugenoltz, 'The New Copyright Directive: Text and Data Mining (Articles 3 and 4)', Kluwer Copyright Blog, 24 July 2019, consultado el 10 Octubre 2024, <https://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/>; Lucchi, 'ChatGPT'; Thomas Margoni and Martin Kretschmer, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology', *GRUR International* 71, no. 8 (1 August 2022): 685–701, <https://doi.org/10.1093/grurint/ikac054>.

14 Véase al respecto, artículo 53, 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 y por el que se modifican los Reglamentos (CE) n.o 300/2008, (UE) n.o 167/2013, (UE) n.o 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial), OJ L, 2024/1689, 12.7.2024.

IA a nivel nacional, como en el caso de Italia, proporcionando, de tal modo, una perspectiva sistemática e interdisciplinar sobre el asunto. Por último, el artículo delinea conclusiones (sección 4).

Es importante señalar que este artículo se dedicará a examinar si la aplicación práctica de la excepción de minería de datos, dentro de los límites establecidos por el marco legal propuesto por el legislador europeo y adoptado o incorporado por los estados miembros de la Unión Europea, desempeña un papel facilitador u obstaculizador en el contexto de la inteligencia artificial generativa. Se abordará el nivel de certeza legal alcanzado, especialmente en lo que respecta a la interacción entre las diferentes soluciones adoptadas a nivel nacional y entre la excepción de minería de datos y el Reglamento Europeo sobre de la inteligencia artificial. En cambio, no es objetivo de este artículo examinar si la opción legislativa elegida por el legislador europeo es adecuada para equilibrar los intereses de los autores de los contenidos protegidos. Por ejemplo, no se discutirá si debería haberse optado por un derecho de remuneración o imponer licencias obligatorias extendidas para el uso de contenido protegido en el entrenamiento de inteligencia artificial.¹⁵ Si por un lado, medidas de protección excesivamente estrictas podrían “obstruir” o “desincentivar” el desarrollo de nuevas herramientas de IA, por el otro, un enfoque laxo hacia la aplicación de los derechos de propiedad intelectual podría socavar la sostenibilidad de los sectores creativos, que dependen en gran medida del valor y el reconocimiento de las creaciones intelectuales. De allí que los efectos del panorama internacional en su conjunto y las opciones de política legislativas disponibles constituye un paso fundamental para involucrarse en discusiones internacionales más profundas sobre posibles futuras vías regulatorias (incluso internacionales), que debería ser abordado en futuras investigaciones.¹⁶

Asimismo, este artículo no tiene como objetivo realizar un análisis sistemático del conjunto de excepciones y limitaciones al derecho de autor que podrían desempeñar un papel relevante, ya sea como obstáculos o facilitadores, en el entrenamiento de la inteligencia artificial.¹⁷

15 Para un enfoque sobre el asunto, véase: Benjamin Raue, ‘Die Freistellung von Datenanalysen durch die neuen Text und Data Mining-Schranken (§§44b, 60d UrhG)’, *Zeitschrift für Urheber- und Medienrecht (ZUM)* 10 (2021): 793–802; Jonathan Pukas, ‘KI-Trainingsdaten und erweiterte kollektive Lizenzen’, *Gewerblicher Rechtsschutz und Urheberrecht* 9 (2023): 614.

16 En el mismo sentido, véase: Jan Bernd Nordemann and Jonathan Pukas, ‘Copyright Exceptions for AI Training Data—Will There Be an International Level Playing Field?’, *Journal of Intellectual Property Law & Practice* 17, no. 12 (1 December 2022): 973–74, <https://doi.org/10.1093/jiplp/jpac106>.

17 Para un análisis sistemático al respecto, véanse los siguientes estudios realizados en coautoría: Caterina Sganga et al., ‘Copyright Flexibilities: Mapping and Comparative Assessment of Eu and National Sources’, SSRN Scholarly Paper (Rochester, NY, 16

2. LA EXCEPCIÓN DE MINERÍAS DE DATOS EN EL DERECHO EUROPEO: ¿PUEDE FAVORECER EL DESARROLLO DE LA IA?

En una economía digital, el valor de los datos y el texto no radica en su aislamiento, sino en la extracción de ideas a partir del análisis colectivo de los mismos. Una tarea como esta sería muy difícil de completar manualmente, de ahí que la utilidad de la minería de datos radica en la posibilidad de analizar, de manera automatizada, grandes corpus de texto y datos, como publicaciones científicas o conjuntos de datos de investigación, para descubrir patrones y nuevas relaciones.¹⁸ Como antes señalado, dicha técnica puede emplearse para 'entrenar' algoritmos de inteligencia artificial (IA) y

January 2023), <https://doi.org/10.2139/ssrn.4325376>; Directorate-General for Research and Innovation (European Commission) et al., 'Improving Access to and Reuse of Research Results, Publications and Data for Scientific Purposes: Study to Evaluate the Effects of the EU Copyright Framework on Research and the Effects of Potential Interventions and to Identify and Present Relevant Provisions for Research in EU Data and Digital Legislation, with a Focus on Rights and Obligations' (Publications Office of the European Union, 2024), <https://data.europa.eu/doi/10.2777/633395>.

- 18 Existe un amplio consenso sobre el potencial de la minería de datos en diversos campos: las herramientas de procesamiento y gestión de datos se utilizan ampliamente en varias industrias, incluidas las de medios de comunicación, farmacéutica, médica, investigación científica y servicios de internet para el consumidor. Así las cosas, se ha evidenciado que la minería de datos es beneficiosa por varias razones. Primero, permite a los investigadores analizar y extraer información valiosa de grandes volúmenes de texto y datos que sería imposible de procesar manualmente. Les permite identificar patrones, tendencias y correlaciones que pueden contribuir a descubrimientos científicos, innovación y avances en diversos campos. Segundo, promueve la investigación interdisciplinaria al permitir que los investigadores de diferentes disciplinas accedan y analicen conjuntos de datos diversos. Esta interacción de ideas y conocimientos puede conducir a nuevas perspectivas y avances en la investigación. Tercero, facilita la exploración del conocimiento existente y la generación de nuevo conocimiento. Al explorar grandes cantidades de texto y datos, los investigadores pueden descubrir conexiones ocultas, validar teorías existentes y generar nuevas hipótesis para una mayor investigación. Cuarto, puede potencialmente mejorar la eficiencia y efectividad de los procesos de investigación. Puede automatizar tareas repetitivas, acelerar el análisis de datos y reducir errores humanos, lo que permite a los investigadores centrarse en análisis e interpretaciones de mayor nivel. Véase, indicativamente: Eleonora Rosati, *Copyright in the Digital Single Market: Article-by-Article Commentary to the Provisions of Directive 2019/790* (Oxford University Press, 2021), <https://doi.org/10.1093/oso/9780198858591.001.0001>; Christophe Geiger, Giancarlo Frosio, and Oleksandr Bulayenko, 'Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?', *IIC—International Review of Intellectual Property and Competition Law* 49, no. 7 (1 September 2018): 814-44, <https://doi.org/10.1007/s40319-018-0722-2>.

de aprendizaje automático, que dependen en gran medida de métodos de minería y análisis de datos.¹⁹

Desde un punto de vista técnico, la minería de datos implica la creación de copias temporales de datos y cualquier contenido asociado para facilitar su análisis.²⁰ Sin embargo, desde una perspectiva de derechos de autor, este proceso constituye un acto de reproducción y, como tal, queda sujeto a la autorización de su titular.²¹ Es por ello que las relaciones entre derecho de autor e IA son cruciales, ya que determinan cómo se pueden utilizar obras protegidas por derechos de autor en el entrenamiento y funcionamiento de los sistemas de IA. En tal sentido, se ha puesto en evidencia que para permitir el aprendizaje automático en los sistemas de IA se necesitan excepciones robustas para la minería de texto y datos que permitan a dichos sistemas replicar, almacenar y procesar datos de entrenamiento para proponer nuevas soluciones.²²

2.1. LA EXCEPCIÓN DE MINERÍA DE DATOS EN LA DIRECTIVA EUROPEA: OBJETIVOS Y CARACTERÍSTICAS

A los fines promover la investigación científica, la innovación y el avance de la inteligencia artificial en la UE, con la adopción de la Directiva 2019/7903 sobre derechos de autor y derechos conexos en el Mercado Único Digital (DDAMUD), la Unión Europea introdujo dos nuevas excepciones, de

- 19 Margoni and Kretschmer, 'A Deeper Look into the EU Text and Data Mining Exceptions'; Eleonora Rosati, 'Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity', *Asia Pacific Law Review* 27, no. 2 (3 July 2019): 198–217, <https://doi.org/10.1080/10192557.2019.1705525>; Séverine Dusollier, 'The 2019 Directive on Copyright in the Digital Single Market: Some Progress, a Few Bad Choices, and an Overall Failed Ambition', *Common Market Law Review* 57, no. 4 (1 August 2020), <https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2020714.pdf>.
- 20 Véase: Rosati, 'Copyright as an Obstacle or an Enabler?' (describiendo en el detalle el funcionamiento de la minería de datos). Véase también: Peter Brettschneider, 'Text Und Data-Mining – Juristische Fallstricke Und Bibliothekarische Handlungsfelder', *Bibliotheksdienst* 55, no. 2 (1 February 2021): 104–26, <https://doi.org/10.1515/bd-2021-0020>. (Describiendo el funcionamiento del proceso de minería de datos llevado a cabo por algunas universidades en Alemania).
- 21 Artículo 2 'Directiva 2001/29/CE del Parlamento Europeo y del Consejo de 22 de mayo de 2001 sobre la armonización de ciertos aspectos del derecho de autor y derechos afines en la sociedad de la información, OJ L167/10 (InfoSoc).
- 22 Gina Maria Ziája, '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 & Practice* 19, no. 5 (1 May 2024): 453–59, <https://doi.org/10.1093/jiplp/jpae025>.

implementación obligatoria, que permiten el uso de obras protegidas para actividades de minería de datos, bajo ciertas condiciones.²³

Dicha opción legislativa resultaba imprescindible por varias razones. En primer lugar, debido a que la excepción para actos temporales de reproducción, como se establece en el Artículo 5(2)(a) Infosoc, es limitada y podría abarcar solo parcialmente las actividades de minerías de datos, especialmente aquellas económicamente insignificantes.²⁴ Aunque las reproducciones de datos en el contexto del entrenamiento de la IA a menudo son de naturaleza temporal, éstas no son solo componentes integrales subordinados de un procedimiento técnico, sino que tienen una importancia técnica y económica propia debido a su impacto en la calidad del modelo de IA. En segundo lugar, Es importante señalar que, si bien el legislador europeo no tuvo en mente la inteligencia artificial al momento de redactar la propuesta de directiva, hacer que la actividad de extracción de datos y contenido necesaria para el entrenamiento de la IA dependa de la obtención previa de autorización por parte del titular de los derechos de autor implicaría costos transaccionales elevados y plazos incompatibles con los del desarrollo de sistemas de IA.²⁵

Así, con la introducción de la excepción de minería de datos – de implementación obligatoria para los estados miembros de la UE – el legislador europeo buscaba lograr una mayor certeza legal para la utilización transfronteriza de contenido, específicamente con fines de investigación y educación.²⁶ El legislador europeo también buscaba eliminar el riesgo de que las

23 Artículos 3 y 4 DDAMUD. Cabe señalar que la propuesta originaria de la Comisión europea preveía la sola implementación de una excepción de minería de datos con fines de investigación científica. La revisión del texto del Consejo propuso introducir una excepción, sin limitación en los fines ni en los destinatarios beneficiarios de la excepción, aplicable siempre que el uso de obras y otros objetos protegidos mencionados en ella no haya sido expresamente reservado por sus titulares de derechos. Así, se incluyó la excepción del Artículo 4 en el texto final de la Directiva. Véase: Eleonora Rosati, *Copyright in the Digital Single Market: Article-by-Article Commentary to the Provisions of Directive 2019/790* (Oxford University Press, 2021), <https://doi.org/10.1093/oso/9780198858591.001.0001>.

24 De Wolf & Partners et al., *Study on the Legal Framework of Text and Data Mining (TDM)* (Publications Office of the European Union, 2014), <https://data.europa.eu/doi/10.2780/1475>; Rossana Ducato and Alain Strowel, 'Ensuring Text and Data Mining: Remaining Issues with the EU Copyright Exceptions and Possible Ways Out', *European Intellectual Property Review* 43, no. 5 (21 April 2021): 322–37.

25 De Wolf & Partners et al., *Study on the Legal Framework of Text and Data Mining (TDM)*. Además, en el caso de obras fuera de comercio y cuando una base de datos ha sido creada con financiamiento público, no siempre está claro quién es el titular de los derechos lo que haría aún más compleja la gestión de licencias. Véase: Lucie Guibault and Andreas Wiebe, 'Safe to Be Open', 2013, 165, <https://doi.org/10.17875/gup2013-160>.

26 European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions—A Digital Single Market Strategy for Europe* (2015) /* COM/2015/0192 final */ , §2.4 Véase también:

entidades de la UE elijan realizar su actividad de minería de datos fuera de la Unión o que las empresas con sede fuera de la Unión prefieran no invertir en empresas dentro de la UE, eligiendo países como Japón en donde existe una amplia excepción al respecto.²⁷

La primera de estas excepciones, encapsulada en el artículo 3 DDAMUD, está limitada a las actividades de minería de datos con fines de educación e investigación científica.²⁸ En lo específico, el Artículo 3 DDAMUD obliga a los Estados miembros a proporcionar una excepción o limitación a los derechos exclusivos del autor de la base de datos, los derechos sui generis de extracción y reutilización del creador de la base de datos, el derecho de reproducción según la Directiva InfoSoc, y los derechos exclusivos de los editores de prensa para las reproducciones y extracciones realizadas por organizaciones de investigación e instituciones de patrimonio cultural, con el fin de llevar a cabo minería de datos de obras u otros materiales a los que tengan acceso legítimo, para fines de investigación científica.²⁹

Desde el punto de vista subjetivo, la excepción es establecida en favor de organismos de investigación e instituciones de patrimonio cultural. Según la directiva, los primeros comprenden instituciones cuales las universidades, incluidas sus bibliotecas, institutos de investigación y cualquier otra organización cuyo objetivo principal sea realizar investigación científica o llevar a cabo actividades educativas que también impliquen la realización de investigación científica: (a) sin fines de lucro o reinvertiendo todos los beneficios en su investigación científica; o (b) de acuerdo con una misión de interés público reconocida por un Estado miembro; de tal manera que el acceso a los resultados generados por dicha investigación científica no

Maria Bottis et al., 'Text and Data Mining in the EU "Acquis Communautaire" Tinkering with TDM & Digital Legal Deposit', *Erasmus Law Review* 12, no. 2 (2019): 190–208, <https://doi.org/10.5553/ELR.000140>.

- 27 Véase, Considerando 8 DDAMUD "A medida que la investigación se realiza cada vez más con la ayuda de la tecnología digital, existe el riesgo de que la posición competitiva de la Unión como área de investigación se vea afectada, a menos que se tomen medidas para abordar la incertidumbre legal en torno a la minería de textos y datos". Véase también: Christian Handke, Lucie Guibault, and Joan-Josep Vallbé, 'Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research', SSRN Scholarly Paper (Rochester, NY, 7 June 2015), <https://doi.org/10.2139/ssrn.2608513>; Artha Dermawan, 'Text and Data Mining Exceptions in the Development of Generative AI Models: What the EU Member States Could Learn from the Japanese "Nonenjoyment" Purposes?', *The Journal of World Intellectual Property* 27, no. 1 (2024): 44–68, <https://doi.org/10.1111/jwip.12285>.
- 28 El Artículo 2(2) DDAMUD, define la minería de datos como cualquier técnica analítica automatizada destinada a analizar texto y datos en formato digital para generar información, que incluye pero no se limita a patrones, tendencias y correlaciones.
- 29 El Considerando 12 DDAMUD aclara que el término "investigación científica" abarca tanto las ciencias naturales como las ciencias humanas.

pueda ser disfrutado de manera preferencial por una empresa que ejerza una influencia decisiva sobre dicha organización.³⁰ Por instituciones de patrimonio cultural se entienden bibliotecas, museos accesibles al público, archivos una institución de patrimonio cinematográfico o de audio.³¹

De conformidad con el Artículo 3(2) DDAMUD, los Estados miembros deben permitir que los beneficiarios almacenen copias de obras u otros materiales siempre que dicho almacenamiento se realice con un nivel adecuado de seguridad. Además, es posible conservar las copias y extracciones siempre que sea necesario a fines de la investigación científica, incluida la verificación de resultados de investigación, por ejemplo, en el caso de la revisión por pares científicos.³²

Por último, el Artículo 7(1) DDAMUD sustrae la excepción de la disponibilidad de las partes, al imponer que esta excepción no puede ser anulada por contrato. El Artículo 7(2) DDAMUD dispone que la excepción debe aplicarse en cumplimiento con el *three step test*, prohibiendo, además de restringir (o anular) su alcance mediante el uso de Medidas Tecnológicas de Protección.

La segunda excepción, regulada en el artículo 4 de dicha directiva, merece especial atención debido a su amplitud y, en consecuencia, por su potencial aplicación en el aprendizaje automático con fines de lucro, aspecto esencial para los sistemas de IA. La norma introduce una excepción o limitación a los derechos exclusivos del autor o derechos conexos, los derechos sui generis de extracción y reutilización del creador de la base de datos para las reproducciones y extracciones de obras, base de datos, incluso el derecho exclusivo de reproducir, traducir, adaptar, arreglar y alterar de cualquier otra manera un programa de ordenador, sin limitación alguna desde el punto de vista subjetivo. Es decir, según esta disposición, cualquier individuo, por ejemplo, desarrolladores comerciales de sistemas de IA pueden hacer copias de obras, programas de ordenador o bases de datos con el propósito de extraer información de texto y datos, siempre y cuando se trate de un usuario legítimo, es decir, con acceso legítimo al contenido u obra.

A diferencia del artículo 3 DDAMUD, la excepción consagrada en el Artículo 4 DDAMUD no tiene una limitación de propósito de uso, pudiendo aplicarse también para la minería de datos con fines de lucro. Esto hace a

30 Artículo 2(1) DDAMUD.

31 Artículo 2(3) DDAMUD. Además, favoreciendo una interpretación extensiva, el Considerando 13 DDAMUD sugiere que por instituciones de patrimonio cultural deben también entenderse las bibliotecas y museos accesibles al público, independientemente del tipo de obras u otros materiales que posean en sus colecciones permanentes, así como archivos, instituciones de patrimonio cinematográfico o de audio, bibliotecas nacionales y archivos nacionales, y las bibliotecas accesibles al público de establecimientos educativos, organizaciones de investigación y organizaciones de radiodifusión pública.

32 Considerando 15, DDAMUD.

la excepción particularmente atractiva para los desarrolladores de sistemas de IA, que podrían avalarse en dicha excepción para utilizar contenido protegidos por derechos de autor a los fines del aprendizaje de modelos de IA destinados a utilizarse o incorporarse en aplicaciones o productos con fines comerciales. Sin embargo, el Artículo 4 DDAMUD deja abierta la posibilidad para los titulares de dichos derechos opten por excluir la excepción respecto de sus contenidos,³³ con el inmediato efecto de que, si el titular ha optado por reservar sus derechos, cualquier actividad de minería de datos quedaría sujeta a la autorización del titular. Ello en cuanto el Considerando 18 especifica que el Artículo 4 DDAMUD no debe afectar indebidamente las oportunidades de explotación, por ejemplo, a través de licencias, para los titulares de derechos en relación con los usos de sus obras que quedan fuera del ámbito del Artículo 3 DDAMUD.

Para excluir la excepción, el mismo Artículo 4, párrafo 3 DDAMUD requiere que los titulares interesados expresen su reserva "de manera apropiada, como, por ejemplo, por medios legibles por máquina en el caso de contenido puesto a disposición en línea" (incluidos los metadatos y los términos y condiciones de un sitio web o un servicio). En todos los otros casos, la reserva de derechos puede realizarse por otros medios, como, por ejemplo, acuerdos contractuales o una declaración unilateral de su titular.

En lo que respecta a la retención de copias una vez que haya concluido la minería de datos, el Artículo 4(2) DDAMUD es menos permisivo. Pues en este caso, la retención de reproducciones y extracciones solo es posible mientras sea necesario para las actividades de minería de datos; esto se debe a que la funcionalidad de una copia para la minería de texto o datos cesa en el momento en que se completa. Por lo tanto, a diferencia del artículo 3 DDAMUD (usos científicos), las copias no pueden ser retenidas para fines más allá de los de la minería de datos, como para verificar y demostrar logros. Así, en el ámbito del entrenamiento de la IA, la retención de dichos resultados estaría justificada solo durante la fase de entrenamiento, quedando excluidos cualquier forma de uso adicional necesario para el uso del texto y los datos extraídos como resultado de su análisis computacional, es decir, el uso que los sistemas de IA podrían potencialmente hacer de ellos.³⁴

33 Mientras que el Artículo 7(1) DDAMUD, no se aplica a esta disposición, lo que la hace modificable por contrato, el Artículo 7(2) DDAMUD requiere que esta excepción esté subordinada a la prueba de tres pasos consagrada en el Artículo 5(5) de la InfoSoc, y que debe preservarse contra la operación de TPMs según lo previsto en el Artículo 6(4) de la InfoSoc.

34 Eleonora Rosati, 'No Step-Free Copyright Exceptions: The Role of the Three-Step in Defining Permitted Uses of Protected Content (Including TDM for AI-Training Purposes)', SSRN Scholarly Paper (Rochester, NY, 10 November 2023), <https://doi.org/10.2139/ssrn.4629528>; Kalpana Tyagi, 'Copyright, Text & Data Mining and the

Dicho esto, los desarrolladores que tengan la intención de utilizar obras protegidas por derechos de autor para entrenar un sistema de IA generativo deberán seguir tres pasos. En primer lugar, deberán obtener acceso legítimo a los datos. En segundo lugar, deberán verificar que los titulares de los derechos no hayan reservado el derecho de hacer reproducciones para fines de minería de datos. Por último, deberán mantener las copias realizadas solo durante el tiempo necesario para fines de minería de datos.

Con la incorporación del artículo 4, el legislador europeo buscaba incentivar la innovación y proporcionar una mayor certeza legal para las actividades de TDM que quedan fuera del ámbito del Artículo DDAMUD, en particular adoptando un escenario estandarizado para los desarrolladores de sistemas en toda la UE, permitiéndoles llevar a cabo actividades de minería de datos de manera legal.³⁵ Si embargo, "buenas intenciones no siempre llevan buenos resultados" y, en lo específico, cabe cuestionar si el marco legal de las excepciones de TDM resulta útil en la práctica, dada la existencia de varias incertidumbres sobre el alcance de la excepción, sus requisitos y la variedad de soluciones nacionales en sede de implementación de la directiva. Las siguientes secciones abordarán este interrogante, analizando, respectivamente, los límites de la formulación del legislador europeo y la fragmentación en la implementación a nivel nacional.

2.2. LÍMITES DE LA NORMA: ¿FACILITADOR U OBSTACULIZADOR?

Varios académicos han llegado a la conclusión de que, a pesar de la introducción de la excepción para la minería de datos en la DDAMUD, su formulación es insuficiente, pudiendo los derechos de autor potencialmente pueden restringir sustancialmente la posibilidad de llevar a cabo actividades de minería de datos con fines de entrenamiento de la IA en Europa.³⁶

Innovation Dimension of Generative AI', *Journal of Intellectual Property Law & Practice*, 9 March 2024, jpa028, <https://doi.org/10.1093/jiplp/jpa028>.

35 Eleonora Rosati, *Copyright in the Digital Single Market: Article-by-Article Commentary to the Provisions of Directive 2019/790* (Oxford University Press, 2021), <https://doi.org/10.1093/oso/9780198858591.001.0001>.

36 Pamela Samuelson, 'The EU's Controversial Digital Single Market Directive—Part II: Why the Proposed Mandatory Text- and Data-Mining Exception Is Too Restrictive', *Kluwer Copyright Blog*, 12 July 2018, consultado el 1 Octubre 2024, <https://copyrightblog.kluweriplaw.com/2018/07/12/eus-controversial-digital-single-market-directive-part-ii-proposed-mandatory-text-data-mining-exception-restrictive/>; Directorate-General for Research and Innovation (European Commission) et al., 'Improving Access to and Reuse of Research Results, Publications and Data for Scientific Purposes: Study to Evaluate the Effects of the EU Copyright Framework on Research and the Effects of Potential Interventions and to Identify and Present Relevant Provisions for Research in EU Data and Digital Legislation, with a Focus on Rights and Obligations' (Publications Office of the European Union, 2024), <https://data.europa.eu/>

En cuanto al artículo 3 DDAMUD, uno de los aspectos más criticados de la normativa es la restricción en la gama de beneficiarios que, junto con los límites en cuanto a los fines de uso, tiene el efecto de reducir considerablemente la portada de la excepción.³⁷ Si bien la norma no excluye la aplicabilidad de la excepción de minería de datos a las asociaciones público-privadas, la misma descarta que esto pueda ser posible cuando una empresa comercial tenga una influencia y control decisivos sobre la organización de investigación en cuestión.³⁸ Este *caveat* podría tener un impacto significativo en el avance de los modelos de modelos de IA generativa que se generen en el ámbito de proyectos de cooperación e investigación, que a menudo involucran múltiples partes, sea empresas o *start up* y universidades.³⁹ Pues bien, mientras

doi/10.2777/633395; Artha Dermawan, 'Text and Data Mining Exceptions in the Development of Generative AI Models: What the EU Member States Could Learn from the Japanese "Nonenjoyment" Purposes?', *The Journal of World Intellectual Property* 27, no. 1 (2024): 44–68, <https://doi.org/10.1111/jwip.12285>; Roberto Caso, 'Il Conflitto Tra Diritto d'autore e Ricerca Scientifica Nella Disciplina Del Text and Data Mining Della Direttiva Sul Mercato Unico Digitale', *Il Diritto Industriale* 2 (2020): 118–26; Maria Lillà Montagnani and Giorgio Aime, 'Il Text and Data Mining e Il Diritto d'autore', *Annali Italiani Del Diritto d'Autore Della Cultura e Dello Spettacolo* 26 (2017): 384; Eleonora Rosati, 'Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity', *Asia Pacific Law Review* 27, no. 2 (3 July 2019): 198–217, <https://doi.org/10.1080/10192557.2019.1705525>; Katharina de la Durantaye, "'Garbage In, Garbage Out"—Die Regulierung Generativer KI Durch Urheberrecht', SSRN Scholarly Paper (Rochester, NY, 10 August 2023), <https://papers.ssrn.com/abstract=4571908>; Thomas Margoni and Martin Kretschmer, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology', *GRUR International* 71, no. 8 (1 August 2022): 685–701, <https://doi.org/10.1093/grurint/ikac054>; Caterina Sganga et al., 'Copyright Flexibilities: Mapping and Comparative Assessment of Eu and National Sources', SSRN Scholarly Paper (Rochester, NY, 16 January 2023), <https://doi.org/10.2139/ssrn.4325376>.

37 Caterina Sganga et al., 'Copyright Flexibilities: Mapping and Comparative Assessment of Eu and National Sources', SSRN Scholarly Paper (Rochester, NY, 16 January 2023), <https://doi.org/10.2139/ssrn.4325376>. Sganga et al., 'Copyright Flexibilities'.

38 Eleonora Rosati, 'An EU Text and Data Mining Exception for the Few: Would It Make Sense?', *Journal of Intellectual Property Law & Practice* 13, no. 6 (1 June 2018): 429–30, <https://doi.org/10.1093/jiplp/jpy063>. (Criticando la limitación de la minería de datos a las organizaciones de investigación, entre otros, debido a que no tiene en consideración que no sólo las organizaciones de investigación se dedican a la minería de datos ni tiene en cuenta las realidades de las prácticas de concesión de licencias a los fines de minería de datos.

39 Como puesto en evidencia en un reciente estudio comisionado por la comisión europea, en la práctica actual de la investigación, se tiende a fomentar la colaboración entre investigadores y socios privados. De hecho, algunos programas de financiación europeos y nacionales para la investigación pueden incluso requerir la participación de socios privados y establecer como condición que estos aporten una parte del presupuesto. Véase: Directorate-General for Research and Innovation (European Commission) et al., 'Improving Access to and Reuse of Research Results, Publications and Data for

que las organizaciones de investigación estarían amparadas en la excepción para realizar las actividades de minería de datos, las entidades comerciales, si ejercen un control sustancial, deberían obtener la autorización de los titulares de los derechos, como, por ejemplo, a través de una licencia. Esto conlleva el riesgo de impedir el desarrollo de una gran cantidad de aplicaciones de IA en los mercados posteriores (o, en su término inglés, 'downstream markets'). Así, lejos de proporcionar certeza jurídica, esta situación añade una capa adicional de complejidad al panorama, pudiendo restringir significativamente la capacidad de innovación y desarrollo en el ámbito de la inteligencia artificial, afectando negativamente el avance tecnológico y la competitividad en el mercado. Del mismo modo, un investigador independiente, quedaría fuera del ámbito de la excepción, ya que no pertenece a una organización, aunque la misma actividad podría haberse beneficiado de la excepción si se hubiera realizado, por ejemplo, en una universidad. Por los mismos motivos, tampoco podrían beneficiar de la excepción una empresa o una *start-up* involucrada en minería de datos en el ámbito de un proyecto o actividad investigación científica, sin fines de lucros. Es por ello por lo que algunos autores sugieren que hubiese sido más acertado adoptar la tradicional distinción "con fines de lucro" y "sin fines de lucro".⁴⁰ En definitiva, es crucial que la legislación contemple estos aspectos para no obstaculizar el progreso en el desarrollo de aplicaciones de IA, permitiendo un uso más amplio y flexible de la minería de datos en diversos contextos.

El artículo 4 DDAMUD, aborda un grupo más amplio de posibles beneficiarios y propósitos en comparación con la excepción de la investigación científica. La ampliación del alcance de los beneficiarios más allá de la investigación científica, es decir, para cualquier fin, expande el potencial de la

Scientific Purposes: Study to Evaluate the Effects of the EU Copyright Framework on Research and the Effects of Potential Interventions and to Identify and Present Relevant Provisions for Research in EU Data and Digital Legislation, with a Focus on Rights and Obligations' (Publications Office of the European Union, 2024), <https://data.europa.eu/doi/10.2777/633395>.

- 40 Thomas Margoni and Martin Kretschmer, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology', *GRUR International* 71, no. 8 (1 August 2022): 685–701, <https://doi.org/10.1093/grurint/ikac054>. De hecho, del análisis de los documentos preparatorios a la adopción de la Directiva, se desprende que, en sede de propuesta del texto de la Directiva, se consideró más apropiado adoptar la formulación que distingue (y permite) la minería de datos en favor de entidades de investigación en la convicción que dicha formulación sería menos restrictiva respecto de la tradicional distinción "con" o "sin" fines de lucro. Sin embargo, como Margoni y Kretschmer notan, en realidad, el efecto de la formulación es más restrictivo que una limitación en los usos "con fines" de lucro, en la medida que mientras bajo la condición de "sin fines de lucro", una empresa que actúa con fines de investigación estaría amparada por la excepción, en la formulación actual, ello no sería posible.

minería de datos para una multiplicidad de usos, entre ellos, la investigación comercial, el periodismo de investigación, el suministro de información al consumidor, el análisis estadístico y, para lo que es relevante para el propósito de este artículo, cualquier proceso de 'aprendizaje' de inteligencia artificial que implique minería de texto y datos.⁴¹ Sin embargo, la discrecionalidad otorgada a los titulares de derechos para "reservarse" el derecho a realizar extracciones y reproducciones de sus obras a los fines de la minería de datos podría obstaculizar la aplicación efectiva de misma y, en algunos casos, socavar por completo los objetivos previstos. En consecuencia, la excepción de minería de datos con fines distintos de la investigación se vuelve bastante precaria, especialmente porque la operatividad de esta puede verse frustrada por limitaciones contractuales.⁴² Además, la incertidumbre sobre si existe, o no, de una reserva contractual respecto de obras disponibles en línea (por ejemplo, en forma de robot.txt metadatos), puede tener un efecto desalentador respecto de inversiones, la investigación y el desarrollo en el área.⁴³

A los efectos de la minería de datos con fines de entrenamiento de modelos de IA generativa, tanto el artículo 3 como el artículo 4 DDAMUD comparten una limitación significativa, ya que sólo prevén exenciones para posibles infracciones del derecho de reproducción y no para el derecho de distribución, comunicación al público o adaptación. En el contexto de la investigación (colaborativa) que involucra desarrollos de IA, la falta de inclusión del derecho de puesta a disposición o comunicación al público entre el ámbito objetivo de la excepción conlleva a que investigadores que conduzcan actividad de minería de datos amparados por el artículo 3 DDAMUD terminen por igualmente infringir los derechos de autor si transfieren o ponen a disposición de terceras partes copias de las extracciones y reproducciones

41 Véase: Séverine Dusollier, 'The 2019 Directive on Copyright in the Digital Single Market: Some Progress, a Few Bad Choices, and an Overall Failed Ambition', *Common Market Law Review* 57, no. 4 (1 August 2020), <https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2020714.pdf>.

42 Rossana Ducato and Alain Strowel, 'Ensuring Text and Data Mining: Remaining Issues with the EU Copyright Exceptions and Possible Ways Out', *European Intellectual Property Review* 43, no. 5 (21 April 2021): 322-37; Gina Maria Ziaja, '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 & Practice* 19, no. 5 (1 May 2024): 453-59, <https://doi.org/10.1093/jiplp/jpae025>.

43 Directorate-General for Research and Innovation (European Commission) et al., 'Improving Access to and Reuse of Research Results, Publications and Data for Scientific Purposes: Study to Evaluate the Effects of the EU Copyright Framework on Research and the Effects of Potential Interventions and to Identify and Present Relevant Provisions for Research in EU Data and Digital Legislation, with a Focus on Rights and Obligations' (Publications Office of the European Union, 2024), <https://data.europa.eu/doi/10.2777/633395>. Cabe mencionar que actualmente existe al respecto un fuerte debate sobre cómo debe llevarse a cabo la declaración de opt-out.

realizadas, incluso si tales actos se realizan en el marco de las actividades de investigación. Asimismo, no quedarían cubiertos por la excepción aquellos casos en que un modelo de IA generativa entrenado legalmente reproduce y/o comunica y pone a disposición del público contenido protegido de terceros.⁴⁴ Algunos incluso sostienen que la omisión de referencia a la puesta a disposición impediría poner en práctica la excepción a los fines del entrenamiento de IA, desde el momento que para llevar a cabo estas evaluaciones, la IA debe reproducir los archivos a analizar y transferirlos a un llamado "corpus", incluso si estos archivos de entrenamiento se eliminan inmediatamente después del análisis.⁴⁵ De último, curiosamente, el Artículo 4 DDAMUD exime los actos de reproducción de programas informáticos con el propósito de realizar actividades de minería de datos, mientras que el Artículo 3 DDAMUD, que permite las actividades de minería de datos por organizaciones de investigación con fines de investigación científica, no lo hace. La directiva no ofrece ninguna orientación sobre esta distinción.

Otro aspecto que acomuna ambas excepciones es que, para beneficiar de la excepción, el usuario debe tener acceso legítimo a la obra (o base de datos), lo cual, según los considerandos, incluye acceso abierto, licencia y obras disponibles libremente en línea. Así, realizar actividades de minería de datos fuera de los límites de un acuerdo de licencia daría lugar a una infracción de los derechos de autor.⁴⁶ Algunos autores incluso han puesto de manifiesto el riesgo de que los precios de las licencias puedan aumentar si los editores aplicaran una regalía por la minería de datos, lo que en definitiva podría crear una brecha entre las instituciones de investigación y/o entre los Estados miembros más o menos facultosos en cuanto a la innovación.⁴⁷

44 Kalpana Tyagi, 'Copyright, Text & Data Mining and the Innovation Dimension of Generative AI', *Journal of Intellectual Property Law & Practice*, 9 March 2024, jpa028, <https://doi.org/10.1093/jiplp/jpae028>.

45 Jonathan Pukas, 'KI-Trainingsdaten und erweiterte kollektive Lizenzen', *Gewerblicher Rechtsschutz und Urheberrecht* 9 (2023): 614 ; Katharina de la Durantaye, "'Garbage In, Garbage Out"—Die Regulierung Generativer KI Durch Urheberrecht', SSRN Scholarly Paper (Rochester, NY, 10 August 2023), <https://papers.ssrn.com/abstract=4571908>.

46 Por ejemplo, Rosati ilustró el caso de algunas editoriales académicas como Oxford University Press y Springer, cuyas políticas ya permitían la realización de actividades de minería de datos para uso no comercial sin necesidad de una licencia o suscripción con dichas editoriales. Véase: Eleonora Rosati, 'Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity', *Asia Pacific Law Review* 27, no. 2 (3 July 2019): 198–217, <https://doi.org/10.1080/10192557.2019.1705525>.

47 Christophe Geiger, Giancarlo Frosio, and Oleksandr Bulayenko, 'Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?', *IIC—International Review of Intellectual Property and Competition Law* 49, no. 7 (1 September 2018): 814–44, <https://doi.org/10.1007/s40319-018-0722-2>; Rossana Ducato and Alain Strowel, 'Ensuring Text and Data Mining: Remaining Issues with the EU Copyright

Asimismo, las *start-up* se verían fuertemente penalizadas, debido al aumento significativo en los costos de transacción para obtener permisos individuales y administrar los pagos de regalías.⁴⁸ En lo que a la minería de datos con finalidad de investigación científica respecta (Artículo 3 DDAMUD), es importante destacar que las excepciones generales de investigación contempladas en el *aquis comunitario*⁴⁹ no requieren acceso mediante una suscripción u otro contrato. Dado que el objetivo planteado por el legislador europeo era aquel de proporcionar certeza jurídica, reforzando y adaptando el marco normativo de las excepciones o limitaciones al entrono digital, una técnica legislativa más apropiada hubiese sido aquella de alinear los requisitos de aplicación de la minería de datos al resto de las excepciones con fines de investigación.

Esta sección ha abordado la cuestión de si el marco actual es lo suficientemente robusto para facilitar la minería de datos, en particular en el ámbito de la IA generativa, en la UE. Como se ha podido observar, el marco actual presenta limitaciones, entre ellas, el carácter restrictivo de beneficiarios en el caso de minería de datos con fines de investigación científica, pudiendo incluso excluirse la aplicación en la hipótesis, no tan inusual, de colaboraciones entre instituciones públicas y privadas. La posibilidad de reserva de derechos y la restringida portada objetiva de la excepción, que comprende solamente la mera reproducción o extracción constituyen, además, ulteriores barreras al utilizzo de la minería de datos, no solo a los fines del aprendizaje o entrenamiento de la IA, sino también en caso de cualquier utilización posterior de aquellos datos extraídos como resultado de su análisis computacional

Exceptions and Possible Ways Out', *European Intellectual Property Review* 43, no. 5 (21 April 2021): 322–37.

48 Algunos autores proponen, en alternativa, de introducir una remuneración equitativa única, proporcionada por el proveedor del sistema de IA generativa, distinguiendo incluso en casos de usos de obras a fines de aprendizaje de la IA (input) y en los resultados por ésta generados (outputs). Véase: Giancarlo Frosio, 'Should We Ban Generative AI, Incentivise It or Make It a Medium for Inclusive Creativity?', SSRN Scholarly Paper (Rochester, NY, 31 July 2023), <https://doi.org/10.2139/ssrn.4527461>; Martin Senftleben, 'Generative AI and Author Remuneration', *IIC—International Review of Intellectual Property and Competition Law* 54, no. 10 (1 November 2023): 1535–60, <https://doi.org/10.1007/s40319-023-01399-4>. En cambio, otros autores sugieren de utilizar el modelo de licencias colectivas extendidas. Katharina de la Durantaye, "'Garbage In, Garbage Out"—Die Regulierung Generativer KI Durch Urheberrecht', SSRN Scholarly Paper (Rochester, NY, 10 August 2023), <https://papers.ssrn.com/abstract=4571908>.

49 En lo específico, Artículo 5(3)(a) Infosoc, Artículo 10(1)(d) Directiva 2006/115/CE del Parlamento Europeo y del Consejo, de 12 de diciembre de 2006, sobre los derechos de alquiler y préstamo establece normas sobre los derechos de autor y derechos afines en lo que respecta a los alquileres y préstamos de obras protegidas, *OJ L 376*, 27.12.2006, Artículos 6(2)(b) y 9(b) Directiva 96/9/CE del Parlamento Europeo y del Consejo, de 11 de marzo de 1996, sobre la protección jurídica de las bases de datos *OJ L 77* de 27/03/1996.

– por ejemplo, para mejorar algoritmos existentes e incluso cuando dichas obras se vean reflejadas en el dato de salida u output. La posibilidad de reserva podría, además, propiciar el desarrollo de sistemas de IA sesgados debido a las condiciones de precio y accesibilidad de los datos de entrenamiento, que generan incentivos erróneos. Así, para evitar pagar licencias, podría resultar económicamente atractivo para los desarrolladores con sede en la Unión Europea entrenar sus algoritmos con datos antiguos, menos precisos y sesgados, o importar modelos de IA ya entrenados en el extranjero con datos no verificables.⁵⁰ En este sentido, la legislación sobre derechos de autor y su excepción a la minería de textos y datos podría constituir un obstáculo para la IA en lugar de un facilitador.

2.3. EL PUZZLE DE IMPLEMENTACIONES NACIONALES

Los estados miembros de la Unión Europea tenían hasta el 7 de junio de 2021 para implementar la excepción para la minería de datos según lo requerido por la DDAMUD.⁵¹ Es importante remarcar que, si bien el modelo delineado por el legislador Europeo constituía un piso mínimo de protección, la directiva dejaba amplio margen de libertad para que los legisladores de los Estados miembros adoptaran modelos más amplios,⁵² incluso pudiendo estos mantener las excepciones de minería de datos ya existentes en la medida que no sean más restrictivas respecto del modelo europeo.⁵³ La mayoría de los Estados miembros no cumplió con la fecha límite de implementación.⁵⁴

50 Martin Kretschmer, Thomas Margoni, and Pinar Oruç, 'Copyright Law and the Lifecycle of Machine Learning Models', *IIC—International Review of Intellectual Property and Competition Law* 55, no. 1 (1 January 2024): 110–38, <https://doi.org/10.1007/s40319-023-01419-3>. Véase también: Katharina de la Durantaye, "'Garbage In, Garbage Out"—Die Regulierung Generativer KI Durch Urheberrecht', *SSRN Scholarly Paper* (Rochester, NY, 10 August 2023), <https://papers.ssrn.com/abstract=4571908> (evidenciando que, mientras en los Estados Unidos se discute los efectos en la innovación que los costos que un mercado de licencias comportaría, el legislador europeo decidió *a priori* de habilitar este mercado).

51 Artículo 29 DDAMUD.

52 Artículo 25 DDAMUD.

53 Por ejemplo, la ley de derechos de Autor Alemana (*Urheberrecht Gesetz*) contenía una excepción de derechos de autor, dentro de la Sección 60d, para fines de minería de datos incluso antes de la adopción de la Directiva CDSM. Esta disposición se introdujo en 2018 para proporcionar una flexibilidad en los derechos de autor para la minería de datos con fines de investigación científica. Véase: Louisa Specht, 'Die neue Schrankenregelung für Text und Data Mining und ihre Bedeutung für die Wissenschaft', *Ordnung der Wissenschaft*, no. 4 (2018): 285–90, <https://doi.org/10.17176/20190131-144947-0>.

54 Véase: Comisión Europea, Medidas nacionales de transposición comunicadas por los Estados miembros, referentes a la Directiva EU 2019/790, disponible en línea < <https://>

Polonia fue el último estado en incorporarla en su derecho interno en el mes de septiembre 2024.⁵⁵

Respecto a las soluciones de implementación adoptada por los legisladores de los estados miembros de la Unión Europea, la mayoría de ellas siguen de cerca el modelo de referencia (artículos 3 y 4 DDAMUD), presentando una escasa o nula variación sustancial con respecto al equivalente de la DDAMUD. Entre ellos, las transposiciones nacionales de Luxemburgo, Malta y los Países Bajos han seguido una incorporación literal del texto de la Directiva.⁵⁶

La "fidelidad" al texto del legislador europeo se evidencia, sobre todo, respecto de los beneficiarios de la excepción de minería de datos con fines de investigación,⁵⁷ para la cual la mayoría de los estados miembros definen los conceptos de instituciones de patrimonio cultural y organizaciones de investigación en consonancia con los artículos 2, apartados 1 y 3, DDAMUD.⁵⁸ Irlanda incluso ha introducido la noción mediante incorporación explícita de la DDAMUD.⁵⁹ En el lado opuesto del espectro se encuentran los estados

eur-lex.europa.eu/legal-content/EN/NIM/?uri=CELEX:32019L0790 > (último acceso el 5 Junio 2024).

55 USTAWA z dnia 26 lipca 2024 r. o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi, consultado el 11 de Octubre 2024, <https://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20240001254/O/D20241254.pdf>. Polonia había presentado el proyecto de ley en el mes de Febrero 2024. Véase: Projekt ustawy o zmianie ustawy o prawie autorskim i prawach pokrewnych oraz niektórych innych ustaw, UC6, Projekt z dnia 14 lutego 2024 r.

56 Luxemburgo (artículo 10, apartado 15, punto 16, artículo 10 bis, artículo 35, apartado 2, y artículo 68 de la Ley de 18 de abril de 2001 sobre derechos de autor, derechos afines y bases de datos), Malta (artículos 3, 4 y 8 del Reglamento sobre derechos de autor y derechos afines al mercado único digital de 2021) y Países Bajos, artículos artículo 15, letra n), letra o), Auteurswet, 'AW'.

57 Artículo 3, DDAMUD.

58 Véase, por ejemplo, la definición de instituciones de patrimonio cultural en Italia (artículo 70-ter (3)(4) l.aut.), Croacia (artículo 187, apartado 3, de la NN), Chipre (artículo 24 de la Ley N° 59/1976 y párrafo 1 del artículo 2 de la Ley N° 59/1976); Chequia (39c(b), Zákon 121/2000) limitada a la definición de las organizaciones de investigación, Grecia (artículo 21A(b)(c) de la Ley 2121/1993); Letonia (Sección 1(19)1 y Sección 212(2), Autortiesību Likums, Ar Grozījumiem: 23.03.2023, 'LCA'); Lituania (artículo 2, apartado 31, punto 35, de la Ley de 18 de mayo de 1999 n. VIII-1185); Malta (artículo 3 del Reglamento sobre los derechos de autor y derechos afines en el mercado único digital de 2021); Portugal (artículo 75, apartado 6, letra a), Código do Direito de Autor e dos Direitos Conexos, 'CDA'), limitado a la definición de las organizaciones de investigación; Rumania (artículo 2.1 de la Lege nr. 26 din 29 februarie 2008 privind protejarea patrimoniului cultural imateria, 'RDA'); Eslovenia (artículo 57 ter, apartado 1, punto 2, Zakon o avtorski in sorodnih pravicah, 'ZASP'); Suecia (artículo 15 quater Lag 1960:729, 'URL'), limitado a las organizaciones de investigación.

59 El artículo 2 de la Copyright and Related Rights Act (CRR) «institución de patrimonio cultural» tiene «el significado que se le asigna en la Unión Europea», una referencia

miembros que se refieren al término "organizaciones de investigación" con listas con fines ejemplificativos, o sin ofrecer una definición, allanando así el camino para diferentes interpretaciones judiciales.⁶⁰ En otros estados miembros, las transposiciones adoptan un enfoque más restrictivo, desde el momento que establecen que la excepción de minería de datos beneficia únicamente a las instituciones mencionadas o reconocidas en leyes específicas. Un ejemplo de este enfoque es adoptado en Estonia, Eslovaquia y, parcialmente, en Bulgaria.⁶¹

Por el contrario, una solución más bien extensiva se encuentra en la incorporación alemana (artículo 60d, apartado 3, de la UrhG-G) y austríaca,⁶² cuyos legisladores han ampliado la gama de beneficiarios para incluir a investigadores individuales que participan ocasionalmente en actividades o proyectos de investigación, siempre que persigan objetivos no comerciales. La máxima flexibilidad se encuentra en España, que define a las organizaciones de investigación como "cualquier entidad que lleve a cabo la investigación científica dentro de sus objetivos principales".⁶³ Sólo unos pocos países han incluido expresamente entre los beneficiarios los hospitales de investigación.⁶⁴ Una solución "sui generis" se encuentra, en cambio, en la normativa francesa. Mientras que, *a primera vista*, la lista de beneficiarios se asemeja a la

que se remite al artículo 2, apartado 3, de la DDAMUD.

60 Bélgica (art. XI.191/1, §1er, 2º-5º Code de droit économique, 'CDE'); Finlandia (artículo 13b, par.2, *Tekijänoikeuslaki* 263/2023); Dinamarca (Artículo 11c Ophavsretsloven, LBK).

61 Estonia (artículo 17/1 AutÕS) define las Organizaciones de Investigación de conformidad con su Ley de Organización de la Investigación y el Desarrollo. Sin vincular el concepto a beneficiarios específicos definidos en una ley separada, el enfoque eslovaco también es bastante restrictivo, ya que solo menciona "bibliotecas, archivos, museos, escuelas o depositarios legales con arreglo a un reglamento especial" (artículo 51b, *Zákon č. 71/2022*, 'ZKUASP'). Una solución similar es adoptada por el legislador búlgaro en lo que respecta las instituciones de patrimonio cultural, en la medida que especifica como destinatarios las "instituciones que preservan el patrimonio cinematográfico o de grabaciones sonoras de acuerdo con la Ley del Fondo Nacional de Archivos" (Artículo 26f, *ЗАКОН ЗА АВТОРСКОТО ПРАВО И СРОДНИТЕ МУ ПРАВА*, 'BCA').

62 Sección 42h(1) UrhG (Alemania) y UrhG (Austria). Hungría adopta un enfoque similar (artículo 33/A(2)(2) del SZJT), que abarca a los investigadores individuales que realizan investigaciones en virtud de un contrato, siempre que ninguna persona u organización que tenga influencia sobre el centro de investigación tenga acceso prioritario a los resultados de la investigación científica en el mismo.

63 España (Artículo 66.3 del Real Decreto 24/2021).

64 Croacia (artículo 187, apartado 2, NN). Bulgaria (artículo 26f, BCA).

Del mismo modo, de acuerdo con la definición proporcionada por Hungría para "organizaciones de investigación" en el artículo 33/A(2)(2) del SZJT, esta excepción se aplica a las actividades científicas no comerciales de los hospitales y laboratorios universitarios, siempre que ninguna persona u organización que tenga influencia sobre el centro de investigación tenga acceso prioritario a los resultados de la investigación científica en el mismo.

de la DDAMUD, un examen más detenido de la formulación nacional revela una mayor flexibilidad, ya que permite a los beneficiarios realizar los actos a través de terceros que actúen en su nombre.⁶⁵

En lo que respecta el ámbito objetivo de la norma, es decir, los actos amparados por excepción, la mayoría de los Estados miembros se adhieren plenamente al modelo de referencia del legislador europeo.⁶⁶ Sólo unas pocas soluciones nacionales extienden el alcance objetivo de la excepción, abarcando derechos exclusivos más allá de la "reproducción" o "extracción" del texto de la Directiva. Sin embargo, el grado de esta flexibilidad varía significativamente y, a veces, se ve eclipsado por la introducción de requisitos adicionales. Por ejemplo, en Italia, el artículo 70-ter (1) l.aut. - correspondiente al artículo 3 DDAMUD - abarca también el acto de comunicación al público de las reproducciones y extracciones realizadas. Sin embargo, este acto sólo está permitido si las extracciones y reproducciones resultantes de las obras se expresan en una nueva obra de manera original. Austria y Alemania permiten la distribución y comunicación de las reproducciones y extracciones realizadas, pero dentro de un grupo específicamente definido de personas para su investigación científica conjunta o, a cualquier persona para examinar la calidad de la investigación científica, con exclusión de cualquier forma de distribución o comunicación con fines de lucro.⁶⁷ En España, las dos vertientes de la excepción de minería de datos abarcan la traducción, adaptación, arreglo y otras transformaciones de programas informáticos.⁶⁸ La excepción rumana, que adopta en cambio una postura más rigurosa,

65 Artículo L. 122-5-3-2 (Code de la propriété intellectuelle, 'CPI').2020

66 Bélgica (art. XI.191/1. §7, art. XI.191/2. §1,3, art. XI.217/1 §6 CDE); Bulgaria (artículo 26e(2), BCA); Croacia (artículo 187/1(NN)), Chipre (artículo 24 de la Ley N° 59/1976); Chequia (39c.b, Zákon 121/2000); Estonia (artículo 19/1 AutÕS); Finlandia (artículo 13b, par.2, *Tekijänoikeuslaki* 263/2023); Grecia (párrafo 2 del artículo 21A de la Ley N° 2121/1993); Letonia (Sección 1(19)1 y Sección 212(1), LCA); Lituania (artículo 22, apartado 1, de la Ley de 18 de mayo de 1999 n. VIII-1185), Luxemburgo (artículo 10, apartado 15, artículos 10 bis y 68 de la Ley de 18 de abril de 2001 sobre derecho de autor, derechos afines y bases de datos); Malta (artículo 4 del Reglamento sobre derechos de autor y derechos afines en el mercado único digital de 2021), Países Bajos (artículo 15n AW), Portugal (artículo 75, apartado 2, letra v), del CDA y artículo 10, apartado 1, letra e), del Decreto Ley n.º 122/2000); Eslovaquia (artículo 51b ZKUASP), Eslovenia (artículo 57b ZASP), Suecia (artículo 15c URL).

67 Artículo 42h.2) UrhG-A (Austria) y artículo 60d.3) UrhG-G (Alemania). En el mismo sentido: Hungría (artículo 35/A(3)-SZJT)

68 Previsto únicamente en el artículo 4 Directiva (UE) 2019/790, pero ausente en el artículo 3 de la misma Directiva. Este enfoque flexible, aunque ligeramente más restringido que la solución española, es seguido también por Francia, que también permite la reproducción de programas informáticos (Artículo. L211-3 CPI – correspondiente al artículo 3 DDAMUD).

ofreciendo a los titulares de derechos la opción de restringir el número de copias que pueden crearse.⁶⁹

Asimismo, en lo que a los requisitos de conservación se refiere, la mayoría de los Estados miembros siguen el modelo europeo, limitándose a exigir que las copias se almacenen con un nivel de seguridad adecuado, con excepción de algunos Estados miembros que especifican de las medidas de seguridad que deben adoptarse.⁷⁰ Por último, Croacia y Bulgaria especifican el concepto de acceso lícito como condición previa para la aplicabilidad de la excepción, en el caso búlgaro incluso imponiendo un plazo de 72 horas a los titulares de derechos para proporcionar el acceso necesario a los usuarios (que tienen acceso legal a los materiales) en el caso de minería de datos con fines científicos.⁷¹ Bélgica impone condiciones adicionales de aplicabilidad, como la indicación de la fuente y el nombre del autor, en la medida de lo posible, que podría volver la excepción de difícil aplicación práctica.⁷²

No hay diferencias significativas respecto de la incorporación en las normativas nacionales del artículo 4 DDAMUD, salvo algunos ejemplos aislados. Irlanda especifica que a los titulares de derechos deben expresar reservas a través de los términos y condiciones de un sitio web o servicio.⁷³ Croacia adopta una solución similar, pero declara explícitamente que en el caso de contenidos en línea "el único medio apropiado" para reservar los derechos es mediante un modo legible por máquina que incorpore metadatos y condiciones generales asociadas a las páginas web o los servicios.⁷⁴ La solución eslovena sugiere que las reservas se hagan utilizando métodos legibles por máquina reconocidos internacionalmente y estandarizados que contengan metadatos y condiciones generales de uso para las obras accesibles públicamente en la Web.⁷⁵ El legislador búlgaro también hizo uso del margen de discrecionalidad otorgado por la directiva, requiriendo a los titulares de

69 Artículo 36⁵ RDA, correspondiente al artículo 3 DDAMUD.

70 Irlanda (artículo 53A, apartado 3A, de la CRRA). También faculta a los titulares de derechos a solicitar información sobre los procedimientos de garantía adoptados; Croacia (artículo 187, apartado 7, de la NN).

71 Artículo 187 (5) NN; Artículo Artículo 26k(2) BCA. Véase también: Ana Lazarova, 'The Last in Line: Bulgaria Implements the CDSM Directive', Kluwer Copyright Blog, 27 December 2023, <https://copyrightblog.kluweriplaw.com/2023/12/27/the-last-in-line-bulgaria-implements-the-cdsm-directive/>. Eslovenia adopta una solución similar al afirmar que si el uso de salvaguardias de seguridad obstaculiza la realización de actos permitidos, los titulares de derechos deben conceder a esa persona acceso a las obras en un plazo no superior a 72 horas (artículo 57b(4) ZASP).

72 Bélgica Art. XI.191/1 § 2.

73 Artículo 53B de la CRRA (incorporando el Considerando de la DDAMUD).

74 Artículo 188 (5) NN.

75 57° ZASP.

los derechos interesados de expresar su reserva "mediante medios técnicos reconocibles por el software que realiza el análisis automatizado".⁷⁶

En lo que a la minería de datos con fines científicos respecta, la heterogeneidad en las definiciones atribuidas a los beneficiarios de la excepción de minería de datos, conforme a la normativa vigente en distintos estados miembros de la Unión Europea, conlleva una serie de implicaciones jurídicas y operativas de importancia para los proyectos de investigación en minería de datos con fines de desarrollo en inteligencia artificial. La falta de armonización o consistencia normativa respecto de los beneficiarios incide en la complejidad interpretativa y en la necesidad de un meticuloso escrutinio de los marcos legales locales pertinentes antes de embarcarse en proyectos colaborativos que involucren la minería de datos a los fines de entrenamiento o desarrollo de IA generativa. Las mismas consideraciones pueden hacerse con referencia al ámbito objetivo de la excepción, para el cual no existe un amplio derecho de comunicación/puesta a disposición del público u otros derechos exclusivos relevantes⁷⁷, pudiendo así verse restringido el acceso a datos cruciales para la investigación, dificultar la colaboración internacional y generar desequilibrios en la capacidad de investigación y desarrollo en el campo de la inteligencia artificial.

Sin embargo, a los fines de IA generativa, podría ser más relevante el Artículo 4 DDAMUD, ya que, a diferencia del Artículo 3 DDAMUD, también permite el uso con fines comerciales. Aquí también, las soluciones nacionales adoptadas varían en su amplitud y requisitos operativos. Estos matices dan como resultado un marco fragmentado y complejo para la minería de datos en la Unión Europea que, desde la perspectiva de los desarrolladores de IA, especie si se trata de *start-up*, representan un costo adicional que puede ser demasiado alto.

Interesante es también de notar que, si bien la excepción de minería de datos puede jugar un rol fundamental en el entrenamiento de la IA, ninguna de las implementaciones nacionales ha expresamente establecido (expandiendo) la relación entre la minería de datos, sea en su vertiente a usos científicos que en aquella que permite los usos comerciales, y el uso de

76 Artículo 26e(4) BCA.

77 Martin Kretschmer, Thomas Margoni, and Pinar Oruç, 'Copyright Law and the Lifecycle of Machine Learning Models', *IIC—International Review of Intellectual Property and Competition Law* 55, no. 1 (1 January 2024): 110–38, <https://doi.org/10.1007/s40319-023-01419-3> (Argumentando que Además, los cambios realizados en el material recopilado pueden constituir una "adaptación" dentro del ámbito del derecho exclusivo y que las excepciones para la investigación o la minería de datos, podrían no cubrir todas las actividades de los investigadores y empresas en este campo).

obras, bases de datos y programas de ordenador en el caso entrenamiento, investigación o posteriores usos de aplicaciones de IA generativa.⁷⁸

3. LA MINERÍA DE DATOS Y LA IA GENERATIVA EN EL REGLAMENTO EUROPEO DE LA IA. PROPUESTAS LEGISLATIVAS NACIONALES

Una de las principales dificultades para los titulares de derechos de autor, cuyos contenidos son 'utilizados' a fines del aprendizaje de la IA, es justamente identificar que sus obras han sido utilizadas en dicho modelo. En el caso de los resultados de salida, u *output*, puede ser aún más identificar obras que se asemejan estrechamente a materiales con derechos de autor. Por ejemplo, en *Andersen v. Stability AI Ltd.*, las artistas demandantes (Sarah Andersen, Kelly McKernan y Karla Ortiz) alegaron que Stability AI, Ltd. y Stability AI, Inc. ('Stability AI'), recopilaron miles de millones de imágenes con derechos de autor de fuentes en línea, sin permiso, para entrenar sus modelos generadores de imágenes para producir nuevas imágenes aparentemente sin atribución a los artistas originales que proporcionaron el material de entrenamiento. En su moción para desestimar las alegaciones, los demandados argumentaron que los modelos no copiaban ni almacenaban ninguna imagen. Más bien, explicaron los demandados, que sus modelos solo analizaban las propiedades de las imágenes en línea para generar parámetros que luego se utilizaban para ayudar al modelo a crear imágenes nuevas y únicas a partir de indicaciones de texto, en lugar de reproducir o copiar alguna parte de las imágenes subyacentes utilizadas para el entrenamiento. En noviembre 2023, el tribunal Distrito Norte de California rechazó la demanda, entre otros, porque no emergía de manera clara si Stable Diffusion almacenaba copias comprimidas de obras protegidas por derechos de autor o si, en su lugar,

78 Cabe mencionar que el primer proyecto legislativo polaco de implementación de la Directiva, aprobado en febrero de 2024 (Projekt ustawy o zmianie ustawy o prawie autorskim i prawach pokrewnych oraz niektórych innych ustaw, UC6, Projekt z dnia 14 lutego 2024 r.), excluía explícitamente la posibilidad de recurrir la excepción de minerías de datos para la creación de modelos de IA generativa. Esta exclusión había sido fuertemente criticada sea por añadir complejidad a un marco ya fragmentado y, sobre todo, en cuanto planteaba fuertes dudas sobre si el legislador polaco había excedido los límites de discrecionalidad establecidos por la Directiva europea. Véase: En este sentido, véase: Paul Keller, 'TDM: Poland Challenges the Rule of EU Copyright Law', *Kluwer Copyright Blog* (blog), 20 February 2024, consultado el 29 Marzo 2024, <https://copyrightblog.kluweriplaw.com/2024/02/20/tdm-poland-challenges-the-rule-of-eu-copyright-law/>. El gobierno polaco dio marcha atrás y la ley final implementa un texto de la excepción que sigue de cerca el texto original de los artículos 3 y 4 DDAMUID.

utilizaba “métodos estadísticos y matemáticos” para capturar y almacenar conceptos de las imágenes de entrenamiento.⁷⁹

Con argumentaciones similares, en *Kadrey v. Meta*,⁸⁰ siempre el tribunal de Distrito Norte de California también rechazó la demanda por infracción de derechos de autor, promovida por un grupo de escritores en contra de Meta, alegando que Meta había reproducido las obras (libros) de los artistas demandantes para entrenar la plataforma de IA, LLaMa (Large Language Model Meta AI). En este caso, el tribunal, aclaró también que para prevalecer en una demanda de infracción fundada en que los *otputs* o datos de salida de LLaMa son obras derivadas, los autores necesitaban alegar y probar que las salidas u *outputs* “incorporan de alguna forma una porción de los libros de los demandantes”. Además, citando la decisión Andersen, el tribunal determinó que los demandantes tendrían que establecer la existencia de una similitud sustancial entre los datos de salidas u *outputs* y las obras supuestamente infringidas (datos de entrada o *input*).

En la Unión Europea, esto ha llevado, entre otras cosas, a que representantes de autores y de las industrias creativas plantearan preocupaciones sobre cómo – y si fuese posible – verificar si el entrenador o desarrollador del modelo de IA ha respetado los límites de la excepción de minería de datos.⁸¹ Si los titulares de derechos no saben si su contenido se está utilizando o se va a utilizar, la mera posibilidad de optar por la reserva no es un instrumento útil para asegurar los derechos de autor.⁸² Algunos incluso han

79 *Andersen v. Stability AI Ltd*, No. 3:23-cv-00201 (N.D. Cal. 2023).

80 *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-03417-VC (N.D. Cal. 2023).

81 European Writer's Council et al., 'Joint Statement on Generative Artificial Intelligence and the EU AI Act', *EWG-European Writers Council* (blog), 25 April 2024, consultado el 30 de Mayo, https://europeanwriterscouncil.eu/240425_cwos_jointstatement_ai-act/. (Entre otros, llamando al legislador europeo a clarificar sobre la relación – y portada aplicativa– de la excepción de minería de datos y la IA generativa). En las mismas líneas se movía la Resolución del Parlamento Europeo sobre los derechos de propiedad intelectual para el desarrollo de tecnologías de inteligencia artificial, adoptada en el mes de octubre 2020. Con la Resolución, el Parlamento Europeo, reconociendo la necesidad de crear un marco legal que aborde los desafíos planteados por la IA generativa, solicitaba la introducción de un sistema que garantiza la transparencia y responsabilidad en el uso y desarrollo de la IA. Véase: European Parliament. (2020). Resolution of 20 October 2020 on Intellectual Property Rights for the Development of Artificial Intelligence Technologies (2019/2166(INI)).

82 Gina Maria Ziaja, '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 & Practice* 19, no. 5 (1 May 2024): 453–59, <https://doi.org/10.1093/jiplp/jpae025>.

llegado a cuestionarse si el uso para entrenar la IA generativa constituye una nueva forma de uso.⁸³

Dado que una modificación de la DDAMUD en el breve plazo no era una opción posible, tanto el legislador europeo, como los estados nacionales, vieron en la vía regulatoria de la IA un lugar apropiado para delinear ulteriormente las relaciones entre la minería de datos y la IA, incluso las tan solicitadas aclaraciones sobre la operatividad del artículo 4 DDAMUD (minería de datos en general, sin limitación de uso). En tal sentido, el siguiente apartado se centrará en los artículos del Reglamento de Inteligencia Artificial que afectan al derecho de autor y a la excepción objeto de examen.⁸⁴

3.1. EL REGLAMENTO EUROPEA DE LA INTELIGENCIA ARTIFICIAL

Como apenas mencionado, en abril de 2021, la Comisión Europea propuso un marco regulatorio para la inteligencia artificial.⁸⁵ A lo largo del proceso legislativo, la propuesta ha recibido numerosas enmiendas hasta llegar a la aprobación del texto definitivo en el mes de mayo de 2024.⁸⁶ Entre ellas,

83 Andrés Guadamuz, 'A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs', SSRN Scholarly Paper (Rochester, NY, 26 February 2023), <https://doi.org/10.2139/ssrn.4371204>; Jonathan Pukas, 'KI-Trainingsdaten und erweiterte kollektive Lizenzen', *Gewerblicher Rechtsschutz und Urheberrecht* 9 (2023): 614; Katharina de la Durantaye, "'Garbage In, Garbage Out' – Die Regulierung Generativer KI Durch Urheberrecht', SSRN Scholarly Paper (Rochester, NY, 10 August 2023), <https://papers.ssrn.com/abstract=4571908>.

84 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 y por el que se modifican los Reglamentos (CE) n.o 300/2008, (UE) n.o 167/2013, (UE) n.o 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial), OJ L, 2024/1689, 12.7.2024 ('Reglamento IA').

85 Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, Brussels, 21.4.2021 COM(2021) 206 final.

86 Durante el "diálogo tripartito"(*), tanto el Consejo de la Unión Europea como el Parlamento Europeo habían propuesto enmiendas al texto originario de propuesta de la Comisión Europea en el 2022 y en el 2023. Véase: Council of the European Union, General Approach, Dec. 6, 2022 (15698/22); European Parliament, Amendments of 14 June 2023 to the Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (AI Act) and amending certain Union acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), P9_TA(2023)0236. Tras un acuerdo provisional sobre el texto alcanzado entre el Consejo de la UE, el Parlamento Europeo y la Comisión Europea el 8 de diciembre de 2023, y la publicación del texto final de compromiso el 26 de enero de 2024, el Parlamento de la UE aprobó el Reglamento de Inteligencia Artificial de la UE el 13 de marzo de 2024, seguida de la aprobación por el Consejo de la Unión Europea el 21 de mayo de 2024, concluyéndose así el proceso legislativo. El nuevo reglamento entrará en vigor luego de 20

cabe mencionar, que el texto originario de la propuesta no contenía alguna referencia a la minería de datos o, en general, al rol del derecho de autor en el desarrollo de la IA. Dicha relación se puso de relieve cuando entre las enmiendas propuestas por el Parlamento Europeo, se incluyeron referencias a los "datos de entrenamiento "protegidos por la ley de derechos de autor", remarcándose que "el impacto adverso causado por el sistema de IA en los derechos fundamentales protegidos por la carta [...] incluye los derechos de propiedad intelectual". Entre otras cosas, las enmiendas publicadas por el Parlamento incluyeron obligaciones adicionales para los proveedores de modelos básicos y sistemas de IA generativos.⁸⁷

En texto definitivo del Reglamento de la IA vincula explícitamente el uso de obras protegidas por derechos de autor para el entrenamiento de modelos de IA a la excepción de minería de datos introducidas con la DDAMUD.⁸⁸ Así, la interrelación entre el Reglamento IA y los derechos de autor, en lo específico la excepción de minería de datos emerge de manera explícita, poniendo fin al debate sobre la aplicabilidad de esta excepción al entrenamiento de modelos de IA. El Considerando confirma, además, que, si los titulares de los derechos han decidido reservar sus derechos para prevenir la minería de textos y datos, los proveedores de modelos deben obtener su autorización para usar dicho contenido protegido con fines de minería de datos.

Para atender las preocupaciones de los autores, el Reglamento IA aborda la cuestión de cómo verificar si los desarrolladores del modelo de IA han

días de publicación y se aplicará dos años después de su entrada en vigor, con algunas excepciones para disposiciones específicas. Esta es la etapa final aprobó el Reglamento de Inteligencia Artificial de la UE. Véase: European Council of the European Union, 'Artificial Intelligence (AI) Act: Council Gives Final Green Light to the First Worldwide Rules on AI', European Council of the European Union, Consilium Europa, 21 May 2024, <https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/>.

(*) En el marco del procedimiento legislativo ordinario de la Unión Europea, un diálogo tripartito es una negociación interinstitucional informal que reúne a representantes del Parlamento Europeo, el Consejo de la Unión Europea y la Comisión Europea. El objetivo de un diálogo tripartito es llegar a un acuerdo provisional sobre una propuesta legislativa que sea aceptable tanto para el Parlamento como para el Consejo, los legisladores.

87 European Parliament, Amendments of 14 June 2023 to the Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (AI Act) and amending certain Union acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), P9_TA(2023)0236.

88 Considerandos 105-109, 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 y por el que se modifican los Reglamentos (CE) n.o 300/2008, (UE) n.o 167/2013, (UE) n.o 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial), OJ L, 2024/1689, 12.7.2024.

respetado los límites de la excepción de minería de datos. En este sentido, el Considerando 107 del Reglamento IA destaca la importancia de la transparencia para garantizar la responsabilidad y facilitar la aplicación de los derechos de autor. Así, el Reglamento de IA incluye al respecto dos requisitos específicos adicionales que los de aplicaciones de IA deben cumplir:⁸⁹ Según el Art. 53 (1) apartado (c), deben establecer y ejecutar una política de derechos de autor que incluya tecnologías de vanguardia para identificar y cumplir con las posibles reservas de exclusión de los autores de conformidad con el Artículo 4(3) DDAMUD. Además, el Art. 53 (1) apartado (d) del Reglamento de IA obliga a los proveedores a “elaborar y hacer pública un resumen suficientemente detallado sobre el contenido utilizado para el entrenamiento del modelo de IA de propósito general”. Para garantizar un estándar razonable y uniforme, la Oficina de IA, también creada por el reglamento, será la encargada de crear una plantilla sobre como proporcionar dicha información o resumen detallado. Según el Considerando 107 del Reglamento IA, el resumen a proporcionar debería ser comprensible en vez de técnicamente detallado, por ejemplo, enumerando “las principales colecciones o conjuntos de datos que se utilizaron para entrenar el modelo”, para facilitar a los titulares de derechos involucrados que puedan ejercitar sus derechos, incluida la posibilidad de reservar sus derechos conforme a la legislación de la Unión. La Oficina de IA también debería supervisar el cumplimiento de estas obligaciones sin realizar evaluaciones individuales del cumplimiento de derechos de autor para cada pieza de datos de entrenamiento.⁹⁰

De esta manera, el Reglamento IA pareciera ofrecer una forma de operacionalizar la facultad de exclusión de los titulares de derechos sobre la base del Artículo 4 DDAMUD, además de un requisito específico de transparencia en el entrenamiento de la IA, a los fines de equilibrar los intereses de los titulares de derechos de las obras utilizadas en el entrenamiento de la IA

89 Esta obligación está dirigida específicamente a los proveedores de modelos básicos que utilizan IA generativa. A estos efectos, los modelos básicos se definen en el Artículo 3(63)(66) como “un modelo de sistema de IA que se entrena con datos amplios a escala, está diseñado para la generalidad de la salida y puede adaptarse a una amplia gama de tareas distintivas”. Según el Considerando 99, los sistemas de IA generativos son “sistemas de IA específicamente destinados a generar, con diferentes niveles de autonomía, contenido como texto complejo, imágenes, audio o video”. Mientras que Artículo en el del 53 (2) del Reglamento excluye estas obligaciones de documentación para los proveedores de modelos de IA que se distribuyen bajo “licencia libre y de código abierto”, el Considerando 106 aborda la dimensión internacional de la excepción de minería de textos, exigiendo a los proveedores que comercialicen IA en la UE de garantizar el cumplimiento de las obligaciones pertinentes en el Acta de IA. En particular, cualquier proveedor que coloque un modelo GPAI en el mercado de la UE debe cumplir, independientemente de la jurisdicción en la que se realizaron los actos relevantes de derechos de autor.

90 Considerando 108, Reglamento IA.

generativa. Sin embargo, existe un problema de coordinación entre ambas normativas, desde el momento que mientras que la excepción de minería de datos se refiere a un mecanismo de exclusión *ex ante* (es decir, la "minería de datos está permitida siempre que el titular de los derechos no haya reservado expresamente esta posibilidad"), la obligación de transparencia se refiere a contenido que ya se ha utilizado. Por otra parte, la amplitud de la obligación de documentación no es prisa de críticas.⁹¹ Dada la gran cantidad de datos que dichos sistemas utilizan para su entrenamiento, el bajo umbral de originalidad, la fragmentación territorial de los derechos de autor y su propiedad, la ausencia de un requisito de registro para las obras y, en general, la falta de armonización de metadatos de propiedad de los derechos sugiere que tal obligación es, en la práctica, imposible de aplicar.⁹² Sin mencionar los costos elevados para su cumplimiento, que solo algunas empresas más facultosas podrían sostener. Si bien el Considerando 109 Reglamento IA reconoce la necesidad de proporcionalidad en los requisitos de cumplimiento, particularmente para las pymes y startups, con el fin de facilitar las actividades de investigación no comercial, no es claro cómo ello se verá puesto en práctica. Así, no solo se vería exacerbada la concentración del mercado y se inhibiría la innovación por parte de empresas pequeñas y medianas, sino que también podría llevar a una disminución en la calidad de los datos de entrenamiento, ya que las empresas estarían incentivadas a depender más del contenido que no está protegido.

3.2. INICIATIVAS NACIONALES: LAS PROPUESTAS EN ITALIA Y FRANCIA

En paralelo al Reglamento de la IA, algunas propuestas de reglamentar la IA o incluso redefinir la operatividad de la minería de datos en el contexto de la IA e IA generativa comienzan a surgir en algunos estados miembros de la Unión Europea.

Tal es el caso de Italia, que en el mes de abril de 2024 presentó un proyecto de ley que introduce disposiciones regulatorias sobre el uso de sistemas

91 En este sentido, véase: Gina Maria Ziaja, '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 & Practice* 19, no. 5 (1 May 2024): 453–59, <https://doi.org/10.1093/jiplp/jpae025>.

92 En este sentido: João Pedro Quintais, 'Generative AI, Copyright and the AI Act', *Kluwer Copyright Blog*, 9 May 2023, consultado el 30 mayo de 2024, <https://copyrightblog.kluweriplaw.com/2023/05/09/generative-ai-copyright-and-the-ai-act/>; de la Durantaye, "'Garbage In, Garbage Out"—Die Regulierung Generativer KI Durch Urheberrecht'. Para una postura contraria, véase: Paul Keller, 'A First Look at the Copyright Relevant Parts in the Final AI Act Compromise', *Kluwer Copyright Blog*, 11 December 2023, consultado el 30 de Mayo 2024 <https://copyrightblog.kluweriplaw.com/2023/12/11/a-first-look-at-the-copyright-relevant-parts-in-the-final-ai-act-compromise/>.

de inteligencia artificial en el sistema legal italiano ("Propuesta Italiana de Ley de IA").⁹³ La relación técnica que acompaña el proyecto de ley deja bien claro que el proyecto de ley no se superpone al reglamento europeo sobre inteligencia artificial, sino que acompaña su marco regulatorio en aquellos espacios propios del derecho interno. Con esta propuesta, el gobierno busca introducir un marco legal para regular el desarrollo y la aplicación de la IA en toda Italia. Al centro de la Propuesta de Ley Italiana de IA está la Gobernanza Nacional de la IA, comprendida la atribución de facultad al gobierno a introducir, mediante decretos legislativos, nuevas normas para la implementación del Reglamento Europeo de IA y definir un marco regulatorio nacional en relación con algunos sectores cruciales en los que el uso de sistemas de IA podría tener un impacto significativo a nivel social y económico (incluidos la investigación, la salud, el empleo, la administración pública, la justicia).

Además, la parte cuarta de la Propuesta Italiana de Ley de IA contiene dos disposiciones específicas sobre derechos de autor y asuntos relacionados con los medios. A los fines de este artículo es relevante el artículo 24 de la Propuesta, que dispone que la reproducción y extracción de obras u otros materiales a través de modelos y sistemas de inteligencia artificial, incluso generativos, están permitidas de conformidad con los artículos 70-ter y 70-quater de la ley de derechos de autor (es decir, remitiéndose a las disposiciones de minería de datos con fines científicos y a la minería de datos en general).⁹⁴ Así siguiendo la línea del principio ya expresado en el artículo 53 co 1 apartado. c del Reglamento Europeo de la IA, la propuesta de Ley Italiana de IA expresamente aclara la aplicabilidad de la minería de datos en el contexto de la IA generativa, fortaleciendo, además, el principio según el cual, salvo en el caso de fines de investigación científica, los titulares de derechos de autor pueden optar por no permitir el uso de su contenido para la minería de textos y datos con fines comerciales. Sin embargo, la propuesta no contiene alguna aclaración sobre cómo debe expresarse la reserva de derechos, aspecto que, como ya se ha podido ver, es uno de los puntos más criticados de la normativa europea. Del mismo modo ausente es al tema de la transparencia y, en particular, a la obligación de documentación de las obras utilizadas para el entrenamiento de la IA.⁹⁵ Es posible que el legislador italiano no haya considerado necesario intervenir en un aspecto ya regulado por el legislador europeo o, más bien, que no esté convencido de la factibilidad de dicha obligación.

93 Disposizioni e delega al Governo in materia di intelligenza artificiale, Disegno di Legge n. 1146, comunicato alla presidenza il 20 Maggio 2024. (Disegno di Legge n.1146/2024).

94 ("Se permite la reproducción y extracción de obras u otros materiales a través de modelos y sistemas de inteligencia artificial, incluidos los generativos, de acuerdo con los artículos 70-ter y 70-quarter"), Artículo 24, Propuesta n. 1146/2024.

95 Correspondiente al artículo 53 (1) apartado (d) del Reglamento de IA.

En el frente francés también se ha entablado una propuesta de Ley al respecto. La propuesta en Francia no está dentro del marco de una reglamentación más amplia de la IA, sino que más bien se trata de un específico proyecto de ley destinado a intervenir directamente sobre la normativa del derecho de autor.⁹⁶ Entre otros aspectos, la propuesta de ley se focaliza en la remuneración de los "autores y artistas que deberían haberse beneficiado de una solicitud de autorización antes de que sus creaciones e interpretaciones sean absorbidas por un algoritmo que las utiliza o se inspira en ellas junto con miles de otras".⁹⁷

La Propuesta establece que "la integración de obras protegidas por derechos de autor en un sistema de software de inteligencia artificial, y consecuentemente su uso, está sujeta a las disposiciones generales del [Código de la Propiedad Intelectual], y por lo tanto requiere autorización del autor o titular de los derechos de propiedad intelectual".⁹⁸ Se dispone que dicha autorización debe obtenerse antes de utilizar cualquier obra protegida. Así, el proyecto un sistema de regalías a cargo de quién incorpora y utiliza, en sistemas de IA, obras intelectuales protegidas por derechos de autor. En la exposición de motivos del proyecto de ley se menciona que esto sería posible mediante la trazabilidad de los autores y artistas cuyo trabajo ha sido utilizado por la IA, y en última instancia, mediante el establecimiento de un sistema de regalías basado en la gestión colectiva. De allí que, el artículo 2 del proyecto atribuye a las sociedades de gestión colectiva la facultad de recolectar las regalías por el uso de dichas obras. Además, en los casos en que se cree una obra artificial a partir de obras de origen no identificado, el artículo 4 de la propuesta de ley dispone se establecería un impuesto destinado a aumentar el valor de la creación en beneficio de las organizaciones de gestión colectiva de derechos de autor. Este impuesto, cuyo importe debería determinarse por decreto, se impondría a la empresa que opera el sistema de IA utilizado para crear dicha obra artificial.

El proyecto de ley tiene el mérito de abordar la remuneración de los autores cuyo trabajo ha sido utilizado por una IA para crear una nueva obra, al introducir un sistema de gestión colectiva. Si bien se trata aún de un proyecto de ley en fase preliminar, en caso de que fuese aprobado, su aplicabilidad práctica plantea varios interrogantes, debido a la falta de coordinación con la ley de derechos de autor, la imprecisión terminológica y la falta de certeza

96 Proposition de loi n°1630 visant à encadrer l'intelligence artificielle par le droit d'auteur du 12 Septembre 2023. Disponible en línea aquí < https://www.assemblee-nationale.fr/dyn/16/textes/l16b1630_proposition-loi >. (Proposition de loi n°1630/2023)

97 Véase, exposición de motivos, Proposition de loi n°1630/2023.

98 Artículo 1, Proposition de loi n°1630/2023.

jurídica.⁹⁹ En particular, la propuesta de ley francesa no proporciona algún tipo de referencia a la relación con las excepciones de minería de datos y cómo garantizar la facultad de los autores de expresar su reserva *ex ante*, entre otros aspectos.

Ambos proyectos deberán seguir sus respectivos cursos legislativos, por lo que queda aún por determinarse si los mismos serán aprobados y, en caso afirmativo, cuál será el texto definitivo de los mismos. Sin embargo, ambas ponen de manifiesto la necesidad de abordar de manera armonizada, equilibrando los intereses de todas las partes en juego, algunos aspectos dejados a la discrecionalidad de los estados miembros, con el fin de evitar que la diversidad de soluciones de los estados miembros socave el objetivo buscado por el legislador europeo de dar una mayor certeza legal para la utilización transfronteriza de contenido.

4. CONCLUSIONES

Este artículo ofrece un análisis sobre las implicaciones y alcance de los artículos 3 y 4 de la Directiva sobre Derechos de Autor en el Mercado Único Digital (DDAMUD) en el contexto de la IA generativa. Como se ha podido observar, aunque la minería de datos posee un potencial considerable en el desarrollo de la IA y su incorporación en el marco legislativo de la UE es crucial, tanto la estructura misma de la excepción como la diversidad de enfoques nacionales no ofrecen un cuadro normativo que permita una aplicación armonizada en la UE, pudiendo limitar su potencial aplicación en el desarrollo de la IA generativa. La restrictiva definición de beneficiarios en la excepción para fines de investigación, la exclusión de derechos cruciales como la comunicación pública, las dificultades operativas relacionadas con la facultad de reserva, y la incertidumbre sobre la calificación de acceso legítimo son aspectos inherentes a la normativa que pueden obstaculizar colaboraciones cruciales para el desarrollo de IA generativa en proyectos conjuntos, o reducir considerablemente su aplicabilidad en este ámbito. Este panorama se ve agravado por la diversidad de enfoques entre los estados miembros de la UE, contribuyendo así a incrementar la incertidumbre existente.

A pesar de que el Reglamento Europeo de IA representa un avance significativo en la intersección entre IA y la minería de datos, su texto exhibe algunas deficiencias de coordinación y falta de claridad sobre cómo deben implementarse ciertas obligaciones. Este escenario, exacerbado por la presencia de propuestas nacionales de regulación, podría intensificar la

99 Véase: Kevin Bercimuelle-Chamot, 'French Copyright Framework for Artificial Intelligence: A Half-Hearted Attempt', *The IPKat* (blog), 16 Octubre 2023, consultado el 10 de Abril 2024, <https://ipkitten.blogspot.com/2023/10/french-copyright-framework-for.html>.

fragmentación existente, complicando sobremanera navegar en el panorama legal para el desarrollo de la IA. Si las obligaciones de la IA con respecto al contenido protegido por derechos de autor utilizado para fines de entrenamiento de IA pueden ser operacionalizadas, y si serán efectivas en mejorar la transparencia de los modelos de IA, tendrá que ser evaluado una vez que la Oficina de IA publique las directrices y modelos relevantes, entre otros aspectos. Así, es imperativo clarificar de manera uniforme los aspectos ambiguos de la directiva en cuando a la amplitud y operatividad de la excepción de minería de datos, a fin de evitar que la diversidad de soluciones de los Estados miembros menoscabe el objetivo del legislador europeo de proporcionar un entorno legal certero para la utilización transfronteriza de contenido y, en última instancia, para maximizar el potencial de la IA.

BIBLIOGRAFIA

- Bercimuelle-Chamot, Kevin. 'French Copyright Framework for Artificial Intelligence: A Half-Hearted Attempt'. *The IPKat* (blog). Accessed 10 June 2024. <https://ipkitten.blogspot.com/2023/10/french-copyright-framework-for.html>.
- Bonadio, Enrico, Plamen Dinev, and Luke McDonagh. 'Can Artificial Intelligence Infringe Copyright? Some Reflections'. In *Research Handbook on Intellectual Property and Artificial Intelligence*, edited by Ryan Abbott, 245–57. Law. Cheltenham, UK: Edward Elgar Publishing, 2022. <https://doi.org/10.4337/9781800881907.00007>.
- Bottis, Maria, Marinos Papadopoulos, Christos Zampakolas, and Paraskevi Ganatsiou. 'Text and Data Mining in the EU "Acquis Communautaire" Tinkering with TDM & Digital Legal Deposit'. *Erasmus Law Review* 12, no. 2 (2019): 190–208. <https://doi.org/10.5553/ELR.000140>.
- Brettschneider, Peter. 'Text Und Data-Mining – Juristische Fallstricke Und Bibliothekarische Handlungsfelder'. *Bibliotheksdienst* 55, no. 2 (1 February 2021): 104–26. <https://doi.org/10.1515/bd-2021-0020>.
- Caso, Roberto. 'Il Conflitto Tra Diritto d'autore e Ricerca Scientifica Nella Disciplina Del Text and Data Mining Della Direttiva Sul Mercato Unico Digitale'. *Il Diritto Industriale* 2 (2020): 118–26.
- Christianson, Paul. 'Billions of Miles of Data: The Autonomous Vehicle Training Conundrum', 30 September 2020. <https://blog.cloudfactory.com/autonomous-vehicle-training-conundrum>.
- Contardi, Magali. '¡IA bajo acusa! ¿el uso de imágenes en el aprendizaje de la inteligencia artificial viola los derechos de autor?: implicancias para la creatividad | LVCENTINVS'. IP Blog. *Lvcentinvs* (blog), 9 March 2023. <https://www.lvcentinvs>.

es/2023/03/09/ia-bajo-acusa-el-uso-de-imagenes-en-el-aprendizaje-de-la-inteligencia-artificial-viola-los-derechos-de-autor-implicancias-para-la-creatividad/.

De Wolf & Partners, Directorate-General for the Internal Market and Services (European Commission) Now known as..., Jérôme de Meeûs d'Argenteuil, Jean-Paul Triaille, and Amélie de Francquen. *Study on the Legal Framework of Text and Data Mining (TDM)*. Publications Office of the European Union, 2014. <https://data.europa.eu/doi/10.2780/1475>.

Dermawan, Artha. 'Text and Data Mining Exceptions in the Development of Generative AI Models: What the EU Member States Could Learn from the Japanese "Nonenjoyment" Purposes?' *The Journal of World Intellectual Property* 27, no. 1 (2024): 44–68. <https://doi.org/10.1111/jwip.12285>.

Directorate-General for Research and Innovation (European Commission), Magali Contardi, Caterina Sganga, Martin Senftleben, Thomas Margoni, Mireille Van Eechoud, Pelin Turan, et al. 'Improving Access to and Reuse of Research Results, Publications and Data for Scientific Purposes: Study to Evaluate the Effects of the EU Copyright Framework on Research and the Effects of Potential Interventions and to Identify and Present Relevant Provisions for Research in EU Data and Digital Legislation, with a Focus on Rights and Obligations'. Publications Office of the European Union, 2024. <https://data.europa.eu/doi/10.2777/633395>.

Ducato, Rossana, and Alain Strowel. 'Ensuring Text and Data Mining: Remaining Issues with the EU Copyright Exceptions and Possible Ways Out'. *European Intellectual Property Review* 43, no. 5 (21 April 2021): 322–37.

Durantaye, Katharina de la. "'Garbage In, Garbage Out"—Die Regulierung Generativer KI Durch Urheberrecht'. SSRN Scholarly Paper. Rochester, NY, 10 August 2023. <https://papers.ssrn.com/abstract=4571908>.

Dusollier, Séverine. 'The 2019 Directive on Copyright in the Digital Single Market: Some Progress, a Few Bad Choices, and an Overall Failed Ambition'. *Common Market Law Review* 57, no. 4 (1 August 2020). <https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2020714.pdf>.

European Council of the European Union. 'Artificial Intelligence (AI) Act: Council Gives Final Green Light to the First Worldwide Rules on AI'. European Council of the European Union. Consilium Europa, 21 May 2024. <https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/>.

European Writer's Council, European Council of Literary Translators' Associations, European Composer and Songwriter Alliance, European Federation of Journalists, European Guild for Artificial Intelligence Regulation, Federation of European Screen Directors, International Federation of Actors, et al. 'Joint Statement on Generative Artificial Intelligence and the EU AI Act'. *EWC—European*

Writers Council (blog), 25 April 2024. https://europeanwriterscouncil.eu/240425_cwos_jointstatement_ai-act/.

Frosio, Giancarlo. 'Should We Ban Generative AI, Incentivise It or Make It a Medium for Inclusive Creativity?' SSRN Scholarly Paper. Rochester, NY, 31 July 2023. <https://doi.org/10.2139/ssrn.4527461>.

Geiger, Christophe, Giancarlo Frosio, and Oleksandr Bulayenko. 'Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?' *IIC—International Review of Intellectual Property and Competition Law* 49, no. 7 (1 September 2018): 814–44. <https://doi.org/10.1007/s40319-018-0722-2>.

Guadamuz, Andrés. 'A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs'. SSRN Scholarly Paper. Rochester, NY, 26 February 2023. <https://doi.org/10.2139/ssrn.4371204>.

Guibault, Lucie, and Andreas Wiebe. 'Safe to Be Open', 2013, 165. <https://doi.org/10.17875/gup2013-160>.

Handke, Christian, Lucie Guibault, and Joan-Josep Vallbé. 'Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research'. SSRN Scholarly Paper. Rochester, NY, 7 June 2015. <https://doi.org/10.2139/ssrn.2608513>.

Hejtmánek, Jan. 'Early Generative AI and Its Impact on Automotive Industry, 2023 Summary'. Deloitte Czech Republic, 2023. <https://www2.deloitte.com/cz/en/pages/technology/articles/early-generative-ai-and-its-impact-on-automotive-industry.html>.

Hugenholtz, P. Bernt. 'The New Copyright Directive: Text and Data Mining (Articles 3 and 4)'. Kluwer Copyright Blog, 24 July 2019. <https://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/>.

Keller, Paul. 'A First Look at the Copyright Relevant Parts in the Final AI Act Compromise'. Kluwer Copyright Blog, 11 December 2023. <https://copyrightblog.kluweriplaw.com/2023/12/11/a-first-look-at-the-copyright-relevant-parts-in-the-final-ai-act-compromise/>.

———. 'TDM: Poland Challenges the Rule of EU Copyright Law'. *Kluwer Copyright Blog* (blog), 20 February 2024. <https://copyrightblog.kluweriplaw.com/2024/02/20/tdm-poland-challenges-the-rule-of-eu-copyright-law/>.

Kretschmer, Martin, Thomas Margoni, and Pinar Oruç. 'Copyright Law and the Lifecycle of Machine Learning Models'. *IIC—International Review of Intellectual Property and Competition Law* 55, no. 1 (1 January 2024): 110–38. <https://doi.org/10.1007/s40319-023-01419-3>.

- Lazarova, Ana. 'The Last in Line: Bulgaria Implements the CDSM Directive'. Kluwer Copyright Blog, 27 December 2023. <https://copyrightblog.kluweriplaw.com/2023/12/27/the-last-in-line-bulgaria-implements-the-cdsm-directive/>.
- Lemley, Mark A. 'How Generative AI Turns Copyright Upside Down'. SSRN Scholarly Paper. Rochester, NY, 21 July 2023. <https://doi.org/10.2139/ssrn.4517702>.
- Lucchi, Nicola. 'ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems'. *European Journal of Risk Regulation*, 29 August 2023, 1-23. <https://doi.org/10.1017/err.2023.59>.
- Manteghi, Maryna. 'Overcoming Barriers to Text and Data Mining in the Era of ChatGPT: The Proposed Data Act as a Game-Changer'. *GRUR International* 73, no. 1 (1 January 2024): 34–44. <https://doi.org/10.1093/grurint/ikad098>.
- Margoni, Thomas, and Martin Kretschmer. 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology'. *GRUR International* 71, no. 8 (1 August 2022): 685–701. <https://doi.org/10.1093/grurint/ikac054>.
- Montagnani, Maria Lillà, and Giorgio Aime. 'Il Text and Data Mining e Il Diritto d'autore'. *Annali Italiani Del Diritto d'Autore Della Cultura e Dello Spettacolo* 26 (2017): 384.
- Nordemann, Jan Bernd, and Jonathan Pukas. 'Copyright Exceptions for AI Training Data—Will There Be an International Level Playing Field?' *Journal of Intellectual Property Law & Practice* 17, no. 12 (1 December 2022): 973–74. <https://doi.org/10.1093/jiplp/jpac106>.
- Pukas, Jonathan. 'KI-Trainingsdaten und erweiterte kollektive Lizenzen'. *Gewerblicher Rechtsschutz und Urheberrecht* 9 (2023): 614.
- Quang, Jenny. 'Does Training AI Violate Copyright Law?' *Berkeley Technology Law Journal* 36 (2021): 1407.
- Quintais, João Pedro. 'Generative AI, Copyright and the AI Act'. Kluwer Copyright Blog, 9 May 2023. <https://copyrightblog.kluweriplaw.com/2023/05/09/generative-ai-copyright-and-the-ai-act/>.
- Quintais, João Pedro, and Sebastian Felix Schwemer. 'The Interplay between the Digital Services Act and Sector Regulation: How Special Is Copyright?' *European Journal of Risk Regulation* 13, no. 2 (June 2022): 191-217. <https://doi.org/10.1017/err.2022.1>.
- Raue, Benjamin. 'Die Freistellung von Datenanalysen durch die neuen Text und Data Mining-Schranken (§§44b, 60d UrhG)'. *Zeitschrift für Urheber-und-Medienrecht (ZUM)* 10 (2021): 793–802.

Rosati, Eleonora. 'An EU Text and Data Mining Exception for the Few: Would It Make Sense?' *Journal of Intellectual Property Law & Practice* 13, no. 6 (1 June 2018): 429–30. <https://doi.org/10.1093/jiplp/jpy063>.

———. 'Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity'. *Asia Pacific Law Review* 27, no. 2 (3 July 2019): 198–217. <https://doi.org/10.1080/10192557.2019.1705525>.

———. *Copyright in the Digital Single Market: Article-by-Article Commentary to the Provisions of Directive 2019/790*. Oxford University Press, 2021. <https://doi.org/10.1093/oso/9780198858591.001.0001>.

———. 'No Step-Free Copyright Exceptions: The Role of the Three-Step in Defining Permitted Uses of Protected Content (Including TDM for AI-Training Purposes)'. SSRN Scholarly Paper. Rochester, NY, 10 November 2023. <https://doi.org/10.2139/ssrn.4629528>.

———. 'The German LAION Decision: A Problematic Understanding of the Scope of the TDM Copyright Exceptions and the Transition from TDM to AI Training'. *The IPKat* (blog). Accessed 15 October 2024. <https://ipkitten.blogspot.com/2024/10/the-german-laion-decision-problematic.html>.

Samuelson, Pamela. 'The EU's Controversial Digital Single Market Directive—Part II: Why the Proposed Mandatory Text- and Data-Mining Exception Is Too Restrictive'. Kluwer Copyright Blog, 12 July 2018. <https://copyrightblog.kluweriplaw.com/2018/07/12/eus-controversial-digital-single-market-directive-part-ii-proposed-mandatory-text-data-mining-exception-restrictive/>.

———. 'Legal Challenges to Generative AI, Part I'. *Communications of the ACM*, 22 June 2023. <https://doi.org/10.1145/3597151>.

Senftleben, Martin. 'Generative AI and Author Remuneration'. *IIC—International Review of Intellectual Property and Competition Law* 54, no. 10 (1 November 2023): 1535–60. <https://doi.org/10.1007/s40319-023-01399-4>.

Sganga, Caterina, Magali Contardi, Pelin Turan, Camilla Signoretta, Giorgia Bucaria, Péter Mezei, and István Harkai. 'Copyright Flexibilities: Mapping and Comparative Assessment of Eu and National Sources'. SSRN Scholarly Paper. Rochester, NY, 16 January 2023. <https://doi.org/10.2139/ssrn.4325376>.

Specht, Louisa. 'Die neue Schrankenregelung für Text und Data Mining und ihre Bedeutung für die Wissenschaft'. *Ordnung der Wissenschaft*, no. 4 (2018): 285–90. <https://doi.org/10.17176/20190131-144947-0>.

Stevens, Morgan. 'Training Musical AI Systems'. *Center for Data Innovation* (blog), 30 June 2023. <https://datainnovation.org/2023/06/training-musical-ai-systems/>.

Strowel, Alain, and Rossana Ducato. 'Artificial Intelligence and Text and Data Mining: A Copyright Carol'. In *The Routledge Handbook of EU Copyright Law*. Routledge, 2021. <https://doi.org/10.4324/9781003156277-19>.

Tyagi, Kalpana. 'Copyright, Text & Data Mining and the Innovation Dimension of Generative AI'. *Journal of Intellectual Property Law & Practice*, 9 March 2024, jpae028. <https://doi.org/10.1093/jiplp/jpae028>.

World Intellectual Property Organization. 'Generative AI: Navigating Intellectual Property'. Geneva, Switzerland: World Intellectual Property Organization (WIPO), 28 February 2024.

———. 'The Frontier of MedTech: AI-Driven Medical Devices'. WIPO, 20 November 2023. https://www.wipo.int/policy/en/news/global_health/2023/news_0021.html.

Ziaja, Gina Maria. '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 & Practice* 19, no. 5 (1 May 2024): 453-59. <https://doi.org/10.1093/jiplp/jpae025>.