

MINERÍA DE TEXTOS Y DATOS E INTELIGENCIA ARTIFICIAL: NUEVAS EXCEPCIONES AL DERECHO DE AUTOR

TEXT AND DATA MINING AND ARTIFICIAL INTELLIGENCE: NEW EXCEPTIONS TO COPYRIGHT

H. Andrés Izquierdo*
Palacio & Izquierdo

There is growing attention and discussion on how national laws or international agreements should allow or prevent the fundamental right to research through text and data mining, and artificial intelligence. This article examines how copyright laws are beginning to regulate the subject matter. The most recent legislative amendments cover the inclusion of new exceptions and limitations in the rules of reproduction, storage, and public communication of works.

The analysis indicates that the standards adopted or used to regulate text and data mining (TDM) vary in their content and purpose from one country to another, which begins to generate a varied, inadequate, and incompatible international environment for project development based in text and data mining and artificial intelligence. The article concludes with some thoughts regarding the right of research, artificial intelligence, data and text mining, and the creation of new limitations and exceptions to copyright law.

KEYWORDS: *Data mining; text mining; artificial intelligence; copyright.*

Existe una creciente atención y discusión sobre cómo las leyes nacionales o los acuerdos internacionales deberían permitir o impedir el derecho fundamental a la investigación por medio de la minería de textos y datos, y la inteligencia artificial. Este artículo examina cómo las leyes de derechos de autor están empezando a regular la materia. Las más recientes modificaciones legislativas comprenden la inclusión de nuevas excepciones y limitaciones en las normas de reproducción, almacenamiento, y comunicación pública de las obras.

El análisis indica que las normas aprobadas o utilizadas para regular la minería de textos y datos (TDM) varían en su contenido y objetivo de un país a otro, lo que empieza a generar un entorno internacional variado, inadecuado e incompatible para el desarrollo de proyectos de investigación basados en minería de textos y datos e inteligencia artificial. Este artículo concluye con reflexiones centradas en el derecho a la investigación, la inteligencia artificial, la minería de textos y datos, y la creación de nuevas limitaciones y excepciones en las leyes de derechos de autor.

PALABRAS CLAVE: *Minería de datos; minería de textos; inteligencia artificial; derechos de autor.*

* Abogado. Master of Laws (LL.M.) en Propiedad Intelectual, Derechos de Autor e Industria del Entretenimiento por la Università degli Studi di Torino. Master of Laws (LL.M.) en Propiedad Intelectual, Derechos de Autor, Derecho de Marcas y Derecho del Entretenimiento estadounidense por la American University Washington College of Law. Asesor legal en Propiedad Intelectual e investigador académico de Inteligencia Artificial, Minería de Textos y Datos, y Derecho Cibernético en American University Washington College of Law. Socio de Palacio & Izquierdo (Washington DC, Estados Unidos). Contacto: info@andresizquierdo.com

Lo expresado en este artículo es a título personal, y no representa la posición de la Universidad o la firma de abogados.

Nota del Editor: El presente artículo fue recibido por el Consejo Ejecutivo de THÉMIS-Revista de Derecho el 27 de enero de 2021, y aceptado por el mismo el 26 de abril de 2021.

I. INTRODUCCIÓN

La inteligencia artificial (en adelante, IA) y el aprendizaje automático o *machine learning* (en adelante, ML) han mostrado un progreso práctico significativo en los últimos años. Los algoritmos de IA y ML están creando una sociedad mucho más automatizada, modificando diversos aspectos de la vida actual. La confluencia única de grandes cantidades de datos (incluida la información y los datos que se pueden encontrar en internet), los nuevos algoritmos y la informática moderna abren un nuevo capítulo para el desarrollo de IA y ML. Los productos básicos y los servicios—incluidas la atención médica, la agricultura o las finanzas— ahora dependen más de los algoritmos de ML¹. Algunos de los ejemplos más destacados de crecimiento del sector privado en la actualidad, como Amazon o Facebook², dependen del uso de ML para optimizar sus métodos de producción, cadenas de suministro, el *marketing* y los precios cobrados a sus consumidores.

Sin embargo, existe un problema interesante: algunos países han aprobado normas para regular el creciente desarrollo en IA al mismo tiempo en el que se establecen estándares internacionales diversos y poco armoniosos para manejar el tema³. A medida que muchas industrias crecen y necesitan más información para adelantar investigación en minería de textos y datos o *text and data mining* (en adelante, TDM), algunas naciones han aprobado leyes para regular esta práctica. Aun así, los resultados evidencian leyes que no son compatibles entre esos mismos países. Por ejemplo, las normas en TDM aprobadas recientemente por la Directiva del mercado único digital (2019) en Europa serán incompatibles en muchos aspectos con la reforma de derechos de autor de TDM japonesa del 2018, o con las recientes sentencias de Google o Hathi-Trust del Tribunal de Apelaciones del Segundo Circuito de los Estados Unidos⁴. La diversa regulación crea un entorno internacional inestable, fragmentado y complejo para los usuarios e investigadores que utilizan información protegida por las leyes de derechos de autor.

Asimismo, la mayoría de los países que aún no inician una actualización de sus leyes en materia de

derechos de autor en TDM solo pueden brindar oportunidades limitadas de poner a disposición obras con derechos de autor. Si los investigadores deciden utilizar las limitaciones y excepciones (en adelante, L&E) existentes —como el derecho a la investigación, la reprografía digital, el uso privado o uso personal—, el no contar con reglas explícitas respecto a TDM se traducirá en altos niveles de incertidumbre con motivo de los nuevos y constantes desarrollos tecnológicos de la IA para explorar la información. Además, los países que cuentan con interpretaciones muy restrictivas en L&E (por ejemplo, Argentina), sin duda harán casi imposible el desarrollo de cualquier proyecto de TDM en que se utilicen copias impresas o digitales de obras protegidas por derechos de autor. Los debates sobre IA y TDM crecen en todo el mundo en diversos escenarios, incluidos la Organización para la Cooperación y el Desarrollo Económicos (en adelante, OCDE) y la Organización Mundial de la Propiedad Intelectual (en adelante, OMPI)⁵. No obstante, a la fecha, no ha surgido ningún modelo de consenso al respecto.

Este artículo adoptará una perspectiva interdisciplinaria. La sección II abordará los conceptos de IA, ML y TDM sobre la base de las investigaciones gubernamentales, científicas y tecnológicas más recientes. La sección III revisará y analizará los aspectos legales con referencia a los contextos regionales y nacionales en TDM y derechos de autor. Se plantearán preguntas sobre la capacidad de las excepciones y limitaciones actuales en las leyes regionales o nacionales para promover el acceso a los datos bajo la perspectiva de los investigadores y los usuarios. La sección IV proporcionará recomendaciones para facilitar el diálogo entre los derechos de autor, la IA y la TDM, al tiempo que mostrará posibles soluciones para futuros cambios de políticas en este campo del derecho.

II. TRANSFORMACIÓN DIGITAL HACIA LA INTELIGENCIA ARTIFICIAL Y EL APRENDIZAJE AUTOMÁTICO

A. Inteligencia artificial

Nadie podría haber imaginado que la pregunta de Alan Turing en 1950 “¿pueden las máquinas pen-

¹ Véase MIT SMR (2020).

² Véase Congressional Research Service (2017); Davis (2020), Metz (2020).

³ Véase Flynn *et al.* (2020); también, Flynn *et al.* (2021). En Europa ver Geiger *et al.* (2018a); Hugenholtz (2019); Quintais (2020); para TDM en los Estados Unidos ver Sag (2019); Sag (2009); Carroll (2019); ver en forma general Okediji (2017); Litman (1994).

⁴ Ver *infra* Sección II.

⁵ Véase Organización Mundial de Propiedad Intelectual (2020).

sar?” sería el futuro del desarrollo tecnológico en la humanidad⁶. Este matemático británico es una de las muchas personas que abrieron el camino a décadas de trabajo sobre la capacidad de una máquina para mostrar un comportamiento inteligente. En la década de 1950, la investigación de la IA se centró, principalmente, en los juegos; en la década de 1960, en algoritmos de búsqueda y resolución de problemas generales (GPS); durante la década de 1970, la IA también exploró la comprensión del lenguaje natural y la representación del conocimiento (Zhongzhi, 2011). No obstante, en la década de 1980, a pesar de que se habían puesto en marcha muchos aspectos diferentes de la IA, su campo aún carecía de éxitos prácticos significativos, una situación que llevó a los conocidos “inviernos de la IA”, en los que disminuyó substancialmente la financiación en Investigación y Desarrollo (MIT Lincoln Laboratory [MIT LL], 2019).

A finales de 1990 y principios del 2000, aquella resurgió con algunos éxitos prácticos. En los últimos años, la IA y el Aprendizaje Automático (ML) han mostrado un progreso práctico significativo. La nueva confluencia de grandes cantidades de datos (incluida la información y los datos que se pueden encontrar en internet), algoritmos y computación moderna está abriendo un nuevo capítulo para la historia de la IA⁷. Las estimaciones indican que el 90% de todos los datos se han creado en los últimos dos años. Hay una mayor capacidad para entrenar los algoritmos existentes con grandes cantidades de muestras de datos. Además, el uso de la computación moderna, particularmente las unidades de procesamiento de gráficos (GPU) y la capacidad de almacenamiento, han permitido que la IA emerja en la última década como no fue posible durante sus primeros 70 años de desarrollo. La combinación de estos elementos ha traído consigo avances dramáticos en IA en redes neuronales o métodos de “aprendizaje profundo” (Congressional Research Service [CRS], 2017)⁸.

Como respuesta a estos avances, varios países y regiones de todo el mundo ya han elaborado informes sobre cómo alistarse para los próximos desafíos. En la carrera por ser pioneros en esta tecnología, la Oficina de Política de Ciencia y Tecnología de la Casa Blanca (OSTP), el Comité de Asuntos Legales del Parlamento Europeo, el Comité de Ciencia y Tecnología de la Cámara de los Comunes en el Reino Unido, el Consejo de Estado chino y el Center for Security and Emerging Technology de la Federación rusa han publicado estrategias para liderar en el campo de la IA. Paralelamente a la acción del gobierno, los comités de expertos también han elaborado diversos documentos de políticas públicas⁹, así como el sector privado y las universidades a través de numerosos estudios y documentos de investigación sobre el tema¹⁰.

1. Inteligencia artificial y *machine learning*

Definir a la IA es de suma importancia, especialmente cuando el tema está pendiente de legislación en todo el mundo. Esta tarea plantea un gran desafío, puesto que se encuentra en una evolución constante, por lo que trae nuevos desarrollos y aplicaciones de forma periódica. Los gobiernos han dado diversas definiciones, algunas amplias, otras más limitadas, pero, en general, no existe una comúnmente aceptada, en parte debido a los diversos enfoques de investigación en el campo¹¹. Esta falta de acuerdo traerá problemas en los futuros procesos legislativos a nivel internacional, regional y nacional. No obstante, la falta de una definición precisa y universalmente aceptada de la IA, probablemente, también ha contribuido a que el campo crezca y avance a un ritmo cada vez más acelerado.

Por otra parte, ML es el nombre de una familia integral de técnicas utilizadas para nuevas formas de análisis de datos en IA (Lehr & Ohm, 2017). Hoy en día existe una mayor capacidad para entrenar los

⁶ Véase Turing (1950).

⁷ Ver definición de Inteligencia artificial en la siguiente sección

⁸ Véase también Davis (2020) y Metz (2020). La Agencia de Proyectos de Investigación Avanzada de Defensa (DARPA) ha identificado tres “olas” de desarrollo de la IA: La primera ola se centró en el conocimiento artesanal que permitió el razonamiento habilitado por el sistema en situaciones limitadas, aunque carecía de la capacidad de aprender o abordar la incertidumbre, como con muchos sistemas basados en reglas desde la década de 1980. La segunda ola, desde aproximadamente la década de 2000 hasta el presente, se ha centrado en los avances en las redes neuronales y el aprendizaje automático (por ejemplo, reconocimiento de imágenes, traducción de idiomas) utilizando modelos estadísticos y conjuntos de grandes datos. La tercera ola se centrará en la adaptación contextual, el aprendizaje y el razonamiento a medida que el sistema encuentra nuevas tareas, avanzando hacia la IA general.

⁹ Ver, por ejemplo, el Grupo de expertos de alto nivel en inteligencia artificial designado por la Comisión Europea, el grupo de expertos en IA en la sociedad de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), o el comité selecto sobre inteligencia artificial del Reino Unido, House of Lords.

¹⁰ Ver, por ejemplo, el Max Planck Institute for Innovation and Competition; Center for International Intellectual Property Studies (CEIPI) en Strasbourg University; PIJIP, American University, Washington College of Law.

¹¹ Para Estados Unidos ver el reporte de la Select Committee on Artificial Intelligence of the National Science & Technology Council (2019); para Europa ver EPRS (2020); para China ver New America (2017).

algoritmos existentes con grandes cantidades de muestras de datos, convirtiéndolos en herramientas esenciales para la predicción y la toma de decisiones (Lehr & Ohm, 2017). Como se precisó anteriormente, con el uso de la informática moderna, en particular las unidades de procesamiento de gráficos (GPU) y la capacidad de almacenamiento, la IA ha crecido mucho respecto a sus primeras décadas de desarrollo.

Actualmente, la IA tiene aplicaciones en sectores como transporte, atención médica, agricultura, finanzas, derecho, ciberseguridad, y defensa¹². Además, muchos productos y servicios, incluidos los filtros de correo no deseado, los diagnósticos médicos, el *marketing* de productos y los vehículos autónomos, dependen ahora de los algoritmos de ML (Coglianese & Lehr, 2017). Algunos de los ejemplos más destacados de crecimiento del sector privado en la actualidad, como Amazon o Zillow, dependen en su utilización para optimizar los métodos de producción, los precios, el *marketing* y las cadenas de suministro (Coglianese & Lehr, 2017, p. 1151). También, el ML ha incursionado en la seguridad nacional y la defensa mundial, una de las principales razones por las que los gobiernos también están buscando avances este campo¹³.

Normalmente, el *machine learning* se analiza desde tres enfoques: aprendizaje supervisado, aprendizaje por refuerzo y aprendizaje no supervisado. En el primero, la máquina ‘aprende’ mediante la supervisión o la enseñanza. Más precisamente, re-

cibe un conjunto de entrenamiento que le permitirá encontrar la respuesta correcta (Carroll, 2019). Bajo este sistema, el científico de datos actúa como un guía para enseñarle al algoritmo las respuestas correctas, pues requiere que “las posibles respuestas del algoritmo ya sean conocidas y que los datos utilizados para entrenar el algoritmo ya estén etiquetados con respuestas correctas” (Castle, 2017) [traducción libre]¹⁴. Este tipo de aprendizaje se encuentra en sistemas diseñados para reconocer objetos (como el reconocimiento facial), sistemas de traducción automática, sistemas de apoyo a la toma de decisiones clínicas (como síntomas y diagnóstico), sistemas para evaluar las solicitudes de préstamos (también para seguros) y sistemas de recomendación de clientes (compras en línea)¹⁵.

El aprendizaje por refuerzo es muy similar, ya que ambos sistemas requieren capacitación a través de ejemplos. No obstante, este último cambia el enfoque del aprendizaje automático del reconocimiento de patrones a la “toma de decisiones secuencial impulsada por la experiencia” (Hassabis *et al.*, 2017, p. 246-247) [traducción libre]¹⁶. Este enfoque de aprendizaje automático generalmente implica la creación de metodologías de capacitación que permitan a un sistema explorar de forma independiente, realizar múltiples pruebas en tareas y aprender de estas experiencias. Como se explicó en una audiencia del Senado de los Estados Unidos en 2016:

[El aprendizaje por refuerzo] explora los desafíos de permitir que los sistemas realicen una

¹² Véase Congressional Research Service (2017). Ver también Davis, (2020), Metz (2020).

¹³ El último enfoque de operaciones cibernéticas mediante el uso de *machine learning* (ML) puede mejorar las “herramientas y capacidades de misión crítica”, ya que los modelos de aprendizaje más recientes pueden detectar cambios de tráfico de distribución, patrones de comportamiento de red anómalos y ataques creativos a medida que se originan (Guerra & Tamburello, 2018). Más precisamente, ML puede ayudar a identificar *malware*, anomalías de red, detección de intrusiones, clasificación de agregación de información, inspección profunda de paquetes y detección de amenazas cibernéticas. En los últimos años, las operaciones cibernéticas reales van desde la piratería en redes gubernamentales o militares hasta ataques cibernéticos a corporaciones y cuentas personales. Los ejemplos más notables incluyen: el incidente de Titan Rain en 2003, cuando las instalaciones del Departamento de Defensa de EE. UU., los laboratorios de la NASA, Lockheed Martin y otros sistemas fueron pirateados y perdieron muchos terabytes de información; los incidentes cibernéticos dirigidos contra Estonia y Georgia, que provocaron graves trastornos en los medios de comunicación, el gobierno y los sistemas bancarios; la lucha de Estados Unidos contra el Estado Islámico de Irak y Siria (en adelante, ISIS); o el ciberataque de 2014 a Sony Pictures Entertainment por parte del grupo Guardianes de la paz, que roba datos y deshabilita sistemas informáticos. Nótese el contraste entre la guerra tradicional, que generalmente dependía de la participación de las fuerzas armadas de un Estado o de un grupo capaz de realizar operaciones militares estándar y los nuevos ciberataques. Ver también Myers (2007).

¹⁴ Texto original: “[...] requires that the algorithm’s possible outputs are already known and that the data used to train the algorithm is already labeled with correct answers”. También puede verse Mokhtarian (2018).

¹⁵ Para dar un ejemplo, un algoritmo binario pudo clasificar 1600 sentencias judiciales de cortes en Estados Unidos en temas de sociedades. En este caso, el algoritmo fue capaz de examinar las palabras de cada sentencia y determinar la probabilidad de sentencia a favor o en contra. Otro ejemplo se puede ver con el algoritmo COMPASS, un software de evaluación de riesgos que se utiliza para pronosticar qué delincuentes tienen más probabilidades de reincidir. Este sistema contiene información sobre delincuentes diseñada específicamente para determinar sus riesgos y necesidades e informar planes dinámicos de casos “que guiarán al delincuente a lo largo de su ciclo de vida en el sistema de justicia penal”. Sin embargo, es relevante mencionar que se ha cuestionado la equidad en este tipo de algoritmos, ya que podría presentar sesgos con los sistemas automatizados de toma de decisiones (ADM). Ver Fagan (2015); Spielkamp (2017); ver también NJCM *et al./Paises Bajos* (2018), sentencia por medio de la cual se le prohibió al gobierno holandés el uso de SyRI, una tecnología utilizada para detectar posibles fraudes a la asistencia social).

¹⁶ Texto original: “sequential decision making driven by experience”.

exploración activa en mundos simulados y reales, que tienen como objetivo dotar a los sistemas con la capacidad de hacer predicciones y realizar acciones físicas con éxito (U.S. Senate, Subcommittee On Space, Science, And Competitiveness, Committee On Commerce, Science, And Transportation, 2016, p. 13) [traducción libre]¹⁷¹⁸.

Un tercer sistema de aprendizaje automático relevante es el aprendizaje no supervisado. A diferencia del aprendizaje supervisado o por refuerzo, el aprendizaje no supervisado funciona sin recibir instrucciones externas, un aprendizaje sin etiquetas de autoría humana. Bajo este sistema, la máquina funciona mediante agrupamiento y reducción de dimensionalidad mientras busca identificar la estructura entre los datos no etiquetados¹⁹. Ejemplos comunes de algoritmos de aprendizaje no supervisados son la investigación de segmentación de audiencia, investigación de clientes, reconocimiento de patrones, o detección de anomalías (operaciones cibernéticas). El aprendizaje no supervisado se caracteriza a menudo como aprendizaje supervisado con un resultado desconocido, ya que muchas veces es difícil generar un criterio objetivo para clasificar los resultados como fallidos o exitosos (Hinton & Sejnowski, 1999). El hecho de que una máquina pueda identificar y resolver problemas de procesamiento de forma independiente nos acerca a la pregunta original de “¿pueden pensar las máquinas?” formulada por Alan Turing en la década de 1950.

2. TDM e IA

La minería de textos y datos se está convirtiendo en una práctica cada vez más común. Se refiere a la capacidad de la IA para interpretar grandes cantidades de datos sin procesar, para, precisamente, ser procesados mediante la identificación de patrones²⁰. Como se ha advertido, la TDM puede ubicarse como aprendizaje por refuerzo o como aprendizaje no supervisado, toda vez que se per-

mite a un sistema explorar de forma independiente, realizar múltiples pruebas en tareas y aprender de estas experiencias, y —algunas veces— la máquina funciona mediante el agrupamiento y reducción de dimensionalidad mientras busca identificar la estructura entre los datos no etiquetados.

Así, a medida que aumenta la escala de la información procesada, la TDM utiliza algoritmos para analizar grandes volúmenes de contenido en busca de patrones e información en períodos de tiempo mucho más reducidos en comparación con un lector humano, al mismo tiempo que aporta nuevos conocimientos y comprensión en diversos campos del saber humano (Geiger *et al.*, 2019, p. 24). Sin que las conexiones se identifiquen o mencionen explícitamente en los artículos, por ejemplo, la TDM se puede utilizar en lingüística para analizar grandes volúmenes de texto y extraer patrones sintácticos o gramaticales (Hugenholtz, 2019); en medicina, para encontrar asociaciones entre un gen y una enfermedad, o entre un fármaco y un evento adverso (Borghini, 2016); y, en general, la TDM es una práctica estándar en la actualidad para diversos temas de investigación —incluida la investigación farmacéutica, periodismo, recuperación de información, e información al consumidor (Hugenholtz, 2019)—. En la siguiente sección, examinaremos el enfoque legal adoptado por diferentes países a medida que las normas TDM se están volviendo desiguales, fragmentadas y complejas para las bibliotecas, los investigadores y los usuarios de la información (Caspers *et al.*, 2017).

III. MINERÍA DE TEXTOS Y DATOS, Y LAS NUEVAS EXCEPCIONES EN DERECHOS DE AUTOR

Como se examinó anteriormente, la minería de textos y datos amplía sustancialmente las capacidades de investigación en la literatura científica y académica. Aun así, para funcionar de manera efectiva, necesita acceso completo a los datos de materiales de texto y bases de datos. Actualmente, existen dos posibilidades para el acceso a estos: la

¹⁷ Texto original:

Research efforts have been underway on the challenges of enabling systems to do active exploration in simulated and real worlds that are aimed at endowing the systems with the ability to make predictions and to perform physical actions successfully.

¹⁸ Bajo este sistema, la máquina aprende a través de recompensas o penalizaciones que están vinculadas a los resultados de tales acciones. Los ejemplos incluyen sistemas de ML que aprenden a jugar un juego mediante recompensas vinculadas a las victorias y con las penalizaciones vinculadas a las derrotas; o sistemas de inversión donde las recompensas están vinculadas a las ganancias financieras, y las sanciones a las pérdidas. Cuando este sistema está funcionando a su máximo potencial, las aplicaciones de IA pueden tomar acciones independientes en el mundo real. Ver también U.S. Senate, Subcommittee On Space, Science, And Competitiveness, Committee On Commerce, Science, And Transportation (2016).

¹⁹ Ver Castaneda *et al.* (2020); ver generalmente Hinton & Sejnowski (1999); Barlow (1989); Le *et al.* (2012).

²⁰ La Directiva 2019/790, en su artículo 2.2, define la minería de textos y datos como: “toda técnica analítica automatizada destinada a analizar textos y datos en formato digital a fin de generar información que incluye, sin carácter exhaustivo, pautas, tendencias o correlaciones” (2019). Véase también Eskevich & Bosch (2016).

primera opción es obtener materiales de acceso abierto que generalmente se encuentran disponibles gratuitamente para su descarga y reutilización; la segunda, es utilizar materiales protegidos por el derecho de autor con reglas de permiso específicas para el acceso y la manipulación.

Bajo la primera opción, comúnmente, los contenidos están disponibles para su descarga bajo una licencia Creative Commons o similar que permite una redistribución y reutilización más abierta que una obra protegida exclusivamente por los derechos de autor. Las publicaciones de acceso abierto van en aumento con bases de datos como PubMed Central con más de 2,75 millones de artículos de texto completo para TDM y otros tipos de reutilización²¹. La segunda opción de acceso a contenidos son las revistas basadas en suscripción, que restringen el acceso a sus publicaciones bajo la legislación de derechos de autor aplicable o a los tipos de contratos suscritos. Algunas colaboraciones entre editores brindan acceso a contenidos basados en suscripción, pero no parecen brindar a los investigadores un acceso completo a la literatura publicada para realizar investigaciones de TDM (Carroll, 2019).

Cuando los derechos de autor son la limitación para el acceso a los contenidos, cada país ostenta un conjunto diferente de reglas para permitir, limitar o, posiblemente, prohibir las actividades de TDM. Algunos países como Japón²², Reino Unido²³ y Alemania²⁴ han desarrollado reglas específicas para el uso de TDM. A nivel regional, la Unión Europea aprobó, en 2019, la Directiva sobre el mercado único digital que incluye explícitamente las actividades de investigación de TDM (Directiva 2019/790, 2019). En Estados Unidos, las cortes han permitido actividades específicas de TDM a través del uso justo de los derechos de autor²⁵. Los países que no han abordado la TDM a través de leyes o fallos judiciales podrían permitir la investigación de TDM por medio de excepciones y limitaciones de derechos de autor: el derecho a la investigación, la reprografía, el uso privado o personal, el almacenamiento digital o las medidas técnicas de protección podrían ser una solución para las investigaciones basadas en TDM. Hay otros ejemplos de países en donde la legislación nacional sobre

derechos de autor cuenta con una interpretación restrictiva que bloquea cualquier posibilidad para la investigación de este tipo²⁶.

La legislación integral sobre TDM se presenta como una herramienta esencial para los investigadores, ya que pueden permitir el acceso legal a materiales y bases de datos, realizar reproducciones, almacenar la información y, en algunos casos, compartir los resultados sin violaciones de derechos de autor (Francoise, 2020). Como se verá en esta sección, los países muestran diversas regulaciones sobre la realización de la TDM.

A. Legislación en minería de textos y datos

Las regulaciones sobre TDM, generalmente, son una excepción al régimen actual de derechos de autor, lo que permite a los usuarios extraer obras a las que tienen acceso lícito. La excepción permite a los investigadores y otros usuarios legítimos realizar actividades del tema sobre grandes cantidades de datos. Esta excepción, usualmente, incluye el acceso a contenidos —en el sitio (como la biblioteca o archivo) y de forma remota— y el derecho a la reproducción de las obras. En otros casos, la excepción también garantiza el derecho a retener copias para su almacenamiento y estudio posterior. En contadas ocasiones, también incluye el derecho de comunicación a fin de poder compartir el *corpus* o los resultados de la investigación con terceras personas, fuera del círculo de trabajo. En las siguientes secciones, se examinarán las limitaciones y excepciones de derechos de autor aprobadas recientemente por la Unión Europea, Reino Unido, Japón y Estados Unidos.

1. Unión Europea

La Unión Europea (en adelante, UE) se encuentra en proceso de modernización respecto a sus normas de derechos de autor para los nuevos y cambiantes entornos digitales (European Commission, 2020b). En 2017, se aprobó un reglamento sobre la portabilidad transfronteriza de los servicios de contenido en línea para permitir a los consumidores que compran o se suscriben a películas, retransmisiones deportivas, música, libros electrónicos y juegos a fin de acceder a los contenidos

²¹ Véase la página web PubMed Central (PMC) (s/f).

²² Véase Ley de Derechos de Autor de Japón (Act 48, 1970).

²³ Véase *Copyright, Designs and Patents Act 1988* (modificado 2017).

²⁴ Véase *Urheberrechtsgesetz* (1965).

²⁵ Véase sentencias judiciales: *Authors Guild, Inc. c. Google, Inc.* (2015) y *Authors Guild, Inc. c. HathiTrust* (2014).

²⁶ Véase Flynn *et al.* (2020). Argentina es un ejemplo de interpretación restrictiva de las excepciones y limitaciones de su ley de derechos de autor. El artículo 10 de la Ley 11.723 de 1933, solo permite reproducir breves fragmentos no superiores a 1000 palabras, lo cual hace casi imposible la investigación por medio de TDM.

mientras viajan por países de la UE. En el mismo año, por medio de una Directiva y un Reglamento, se implementó el Tratado de Marrakech para permitir el acceso de contenidos protegido por derechos de autor a las personas ciegas, con discapacidad visual o con alguna otra discapacidad²⁷. En el 2019, la Directiva del mercado único digital (en adelante, Directiva MUD) adoptó nuevas excepciones y limitaciones de derechos de autor que se centran en la disponibilidad en línea de los contenidos en toda la UE, incluidas las excepciones del artículo 3 y 4 sobre TDM²⁸.

Los artículos 3 y 4 de la Directiva MUD tienen como objetivo proporcionar un entorno armonizado para los investigadores de toda Europa. En primer lugar, la Directiva MUD crea excepciones y limitaciones obligatorias para los investigadores que llevan a cabo proyectos en TDM, lo que garantiza una implementación más uniforme en todo el continente. En segundo lugar, las L&E cubren usos comerciales y no comerciales, generando que la investigación se encuentre disponible para materiales a los que se accede legalmente. Finalmente, las disposiciones contractuales creadas por los titulares de derechos no pueden anular los usos no comerciales garantizados en el artículo (*Geiger et al.*, 2018a). Sin embargo, las dos excepciones no son igualmente sólidas, ya que el uso comercial del artículo 4 puede estar restringido por acuerdos privados (Hugenholtz, 2019), como veremos en la sección siguiente.

a. Artículo 3, Directiva MUD

El artículo 3 de la Directiva MUD (Directiva 2019/790, 2019) crea una excepción de derechos de autor obligatoria para la investigación científica no comercial realizada con TDM. Cualquier disposición contractual contraria a la excepción en TDM será inaplicable y el proyecto en TDM solo puede ser promovido por organismos de investigación e instituciones responsables del patrimonio cultural. La excepción incluye reproducciones, extracciones y almacenamiento de obras; no obstante, el usuario requiere un acceso lícito. Para comprender plenamente el alcance de este artículo, es necesario revisar los beneficiarios de la excepción —organizaciones de investigación, instituciones de patrimonio cultural—; las condiciones de uso, como el acceso lícito y las medidas técnicas

de protección (en adelante, TPM); así como las limitaciones que encontrará la disposición en TDM de la Directiva MUD en comparación con otras legislaciones nacionales.

i. Beneficiarios de la excepción

Según la Directiva MUD, el primer grupo de beneficiarios de la excepción son los organismos de investigación (Directiva 2019/790, 2019). Este tipo de organizaciones cubren una amplia variedad de entidades cuyo objetivo principal es realizar investigación científica o realizarla junto con servicios educativos (Directiva 2019/790, 2019, considerando 12). La descripción de las organizaciones de investigación incluye universidades —y sus bibliotecas—, institutos de investigación o cualquier otra entidad, como hospitales, que realice esta clase de labor. La Directiva MUD limita la organización de investigación exigiendo que trabaje sin fines de lucro o reinvertiendo todos los beneficios en su investigación científica, o trabajando bajo una misión de interés público reconocida por un Estado miembro europeo²⁹. Es necesario tener en cuenta que el considerando 11 de la Directiva MUD permite a las organizaciones de investigación beneficiarse de la excepción cuando sus actividades de investigación se llevan a cabo en el marco de asociaciones público-privadas, pero organizaciones en las que las empresas comerciales tienen una influencia decisiva en el proyecto de investigación —como accionistas o miembros que podrían obtener acceso preferencial a los resultados de la investigación—, no se consideran organizaciones de investigación para tal directiva (Hugenholtz, 2019).

El segundo grupo de beneficiarios de las excepciones de la TDM son las instituciones responsables del patrimonio cultural. La Directiva MUD las define como cualquier biblioteca o museo de acceso público, un archivo o una institución de patrimonio cinematográfico o sonoro (Directiva 2019/790, 2019, considerando 13 & art. 2.3). La definición incluye bibliotecas y museos de acceso público, independientemente del tipo de obras que guarden en sus colecciones y archivos permanentes, películas o audio. La descripción también incluye bibliotecas y archivos nacionales, y cualquier archivo y biblioteca de acceso público de establecimientos educativos, organizaciones de investigación u organizaciones de radiodifusión del sector público.

²⁷ Véase Reglamento 2017/1563 (2017) y Directiva 2017/1564 (2017).

²⁸ Véase Directiva 2019/790 (2019), Directiva 96/9/CE (1996) y Directiva 2001/29/CE (2001). Asimismo, ver de forma general Geiger *et al.* (2018a, pp. 4-15).

²⁹ Dicha norma menciona en su considerando 12 que “dicha misión podrá quedar reflejada, por ejemplo, a través de la financiación pública o de disposiciones de la legislación nacional o de los contratos públicos” (Directiva 2019/790, 2019). Asimismo véase Quintais (2020).

Dado que la definición es tan amplia, algunas instituciones del patrimonio cultural también podrían clasificarse como organizaciones de investigación. Es importante mencionar que la excepción no incluye a personas o investigadores no afiliados, claramente excluidos del MUD³⁰.

ii. Acceso lícito

En cuanto a las condiciones de uso, el acceso lícito es un actor importante en el proceso de TDM. Este incluye el contenido basado en una política de acceso abierto y el obtenido mediante acuerdos contractuales entre los titulares de derechos y los beneficiarios de la excepción (organizaciones de investigación o instituciones de patrimonio cultural) (Directiva 2019/790, 2019, considerando 14). La política de acceso abierto se refiere “[...] a la información científica que es gratuita para el usuario y que es reutilizable” (European Commission, 2020b) [traducción libre]^{31,32}, mientras que el contenido sujeto a un acuerdo contractual cubre todas las suscripciones y el contenido que está disponible en línea. El considerando 14 establece que “en el caso de las suscripciones realizadas por organismos de investigación o instituciones responsables del patrimonio cultural, se considera que las personas vinculadas a ellas y amparadas por esas suscripciones también tienen acceso lícito” (Directiva 2019/790, 2019). Así, el acceso lícito se divide entre servicios gratuitos y servicios pagos.

Sin embargo, las limitaciones al acceso legal no vienen con los contenidos de acceso abierto, sino con los contenidos que requieren un pago por suscripción. Cuando el acceso legal está vinculado a los pagos de los servicios de suscripción, la capacidad presupuestaria de la organización pasa a desempeñar un rol fundamental. Geiger, Frosio y Bulayenko afirman que “solo pocas organizaciones de investigación podrán adquirir licencias para to-

das las bases de datos que sean relevantes para un proyecto de investigación en TDM” (2018c, p. 29) [traducción libre]³³. Como resultado, solo un número limitado de organizaciones de investigación podrá adelantar proyectos de TDM. Si la capacidad de financiamiento de cada institución determina los resultados de la investigación en dicho campo, el acceso lícito aumentará las diferencias entre las instituciones. Además, podría aumentar la brecha científica y de innovación entre las naciones europeas desarrolladas y las menos desarrolladas.

El acceso lícito tiene una capa adicional de control, ya que el artículo 3 inciso 3 permite a los titulares de derechos aplicar medidas técnicas de protección (TPM) “para garantizar la seguridad e integridad de las redes y bases de datos en que estén almacenadas las obras u otras prestaciones. Dichas medidas no irán más allá de lo necesario para lograr ese objetivo” (Directiva 2019/790, 2019)³⁴. El considerando 16 va más allá para explicar el razonamiento detrás de la disposición, ya que enmarca la solicitud en un número potencialmente elevado de solicitudes de acceso y descargas de obras. Por lo tanto, los titulares de derechos pueden hacer uso de medidas técnicas de protección en casos en los que la seguridad e integridad de sus sistemas o bases de datos pueda verse comprometida. No obstante, la Directiva MUD también establece un equilibrio recomendado entre las TPM y los riesgos de seguridad, ya que ambos “no deben exceder de lo necesario para alcanzar el objetivo de garantizar la seguridad e integridad de la red y no han de menoscabar la aplicación efectiva de la excepción” (Directiva 2019/790, 2019, considerando 16).

Sin embargo, el alcance de la excepción del artículo 3 inciso 3 genera varias preguntas, ya que podría ubicar a la UE en una posición de desventaja para la investigación en IA³⁵. Si bien la UE divide el régimen de TDM entre uso no comercial y co-

³⁰ Ver más adelante en la sección del Reino Unido, en la que se incluyen los investigadores no afiliados en la excepción TDM de *la Copyright, Designs and Patents Act* (1988, § 29A).

³¹ Texto original: “[...] access to scientific information that is free of charge to the user and is reusable”.

³² Véanse también los proyectos Open Access Infrastructure for Research in Europe (OpenAIRE), Open Science Policy Platform, Horizon 2020, European Open Science Cloud (EOSC) para programas y grupos de expertos enfocados en el acceso a datos abiertos. Ver, en general, en los Estados Unidos el Open Access Policy de la University of California:

Each Faculty member grants to the University of California a nonexclusive, irrevocable, worldwide license to exercise any and all rights under copyright relating to each of his or her scholarly articles, in any medium, and to authorize others to do the same, for the purpose of making their articles widely and freely available in an open access repository (Office of Scholarly Communication, 2013).

³³ Texto original: “Only few research organisations will be able to acquire licences for all databases that are relevant for a TDM research project”.

³⁴ Asimismo, ver de forma general Dusollier:

Despite the expressed intention of the European lawmaker to safeguard a balance between the rights of the copyright holders and the interests of the users and society at large, the anti-circumvention provisions give the rights owners preference: the protection is broad and surely extends beyond the boundaries of copyright; the exceptions are overridden, albeit the empty promise of the article 6(4) (2003, p. 478).

³⁵ Véase Samuelson (2018), Hilty (2016) y Ducato & Strowel (2019).

mercial, otros países como Estados Unidos y Japón han evitado ese enfoque, permitiendo la TDM sin importar el propósito comercial o el tipo de usuario³⁶. Como se evidenciará en las siguientes secciones, las recientes sentencias judiciales de TDM en los Estados Unidos –*Authors Guild, Inc. c. Google* (2015) y *Authors Guild, Inc. c. HathiTrust* (2014)– han permitido el uso comercial de la investigación de TDM bajo la excepción de uso justo en la ley de derecho de autor. Asimismo, Japón, el primer país en aprobar una disposición de derechos de autor en TDM, no distingue entre uso comercial y no comercial; el análisis de la información no debe realizarse necesariamente con fines de investigación científica, y TDM está abierto a todos los usuarios (*Act 48*, 1970). En consecuencia, dadas las limitaciones entre usos comerciales y no comerciales en la UE, las empresas de este continente podrían enviar al exterior sus proyectos de investigación en TDM, a lugares con una legislación menos restrictiva o, de la misma forma, las empresas fuera del continente podrían evitar el mercado europeo debido a las restricciones de la Directiva MUD (Samuelson, 2018).

b. Artículo 4, Directiva MUD

Contrario al artículo 3, el artículo 4 crea una excepción que no tiene ninguna restricción sobre el tipo de usuario o el propósito comercial; aun así, la excepción puede ser anulada por acuerdos contractuales o una declaración unilateral (Quintais, 2020). Esta excepción de TDM incluye todo tipo de usuarios sin importar su afiliación: cualquier organización, corporación o investigador individual puede hacer uso de la excepción en TDM. Además, la actividad TDM es aplicable a cualquier reproducción y extracción de obras legalmente accesibles, ya sea para uso comercial o no comercial (Directiva 2019/790, 2019, art. 4.1 & 4.2). No obstante, la excepción o limitación es aplicable solo a condición de que los titulares de los derechos no se hayan reservado expresamente el uso de las obras, ya que les da una opción de exclusión voluntaria (Directiva 2019/790, 2019, art. 4.3).

A todos los efectos, el artículo 4 es una disposición que permite a cualquier persona u organización participar en proyectos de TDM siempre que el titular del derecho consienta el uso. El artículo 4 inciso 3 y el considerando 18 dan a los titulares de derechos la opción de reservar sus derechos mediante

“la utilización de medios de lectura mecánica, incluidos los metadatos y las condiciones de un sitio web o un servicio” (Directiva 2019/790, 2019). Sin embargo, también pueden reservarse los derechos por otros medios, “como acuerdos contractuales o una declaración unilateral” (Directiva 2019/790, 2019, art. 4.3). Entonces, en la práctica, un editor puede controlar el acceso que tenga la intención de uso comercial; ostenta la libertad de otorgar licencias mediante acuerdos contractuales, o puede bloquear o prohibir unilateralmente cualquier uso de sus obras. El riesgo surge, como afirma Hugenholtz, en que los editores podrán “ofrecer el acceso pago a TDM, y se mostrarán reacios a otorgar licencias TDM a terceros” (2019). Como resultado, el artículo 4 podría, efectivamente, crear y legitimar un ‘mercado derivado’ para las actividades de TDM, a través de contratos, licencias e incluso, exclusividad sobre obras y contenidos.

Este artículo también podría ubicar a la UE en una posición de desventaja, ya que la opción de exclusión voluntaria en la Directiva MUD no coincide con las disposiciones de TDM menos restrictivas de otras jurisdicciones. A la vanguardia de las regulaciones abiertas de TDM, Japón puede permitir cualquier uso de TDM, incluida la investigación comercial, sin ninguna compensación al titular de los derechos. La legislación de Japón también puede permitir que un usuario de TDM esté libre de la retribución económica de un titular de derechos cuando accede a datos o información, y realiza copias electrónicas incidentales de las obras o durante la búsqueda de bases de datos (*Act 48*, 1970). Además, como se verá en el caso de Google TDM de Estados Unidos, las empresas pueden participar en proyectos de TDM sin ninguna compensación al titular de los derechos si cumplen las condiciones de uso justo requeridas³⁷. Asimismo, algunos países que carecen de disposiciones específicas para limitar el uso de TDM, como Belice o Barbados, pueden permitir la libre explotación en virtud del trato justo o del artículo 9, numeral 2, del Convenio de Berna (1886)³⁸. Estos mercados alternativos podrían alentar a los investigadores, organizaciones y corporaciones a explorar países con menos limitaciones legales para los proyectos de TDM.

2. Reino Unido

La excepción de derechos de autor de TDM del Reino Unido se aprobó en 2014, pero se limita a

³⁶ Ello se comentará en secciones posteriores del presente trabajo.

³⁷ Véase *Copyright Act of 1976* (U.S. Code, 2012, título 17, § 107), Limitaciones a los derechos exclusivos: Uso justo.

³⁸ Véase Ley de Derechos de Autor de Bélize (2000, cap. 252) y Ley de Derechos de Autor de Barbados (1998, cap. 300).

un propósito no comercial³⁹. El texto del artículo 29A del *Copyright, Designs and Patents Act* señala lo siguiente: “una persona que tenga acceso lícito a la obra puede realizar un análisis computacional de todo lo registrado en la obra con el único fin de realizar una investigación con fines no comerciales” (1988, art. 29A(1)(a)) [traducción libre]⁴⁰. El alcance de esta excepción se enmarca en el artículo 5.3 (a) de la Directiva InfoSoc, norma que requiere un fin de investigación científica con carácter no comercial (Directiva 2001/29/CE, 2001)⁴¹. Aunque el artículo 5.3 no es una excepción obligatoria para los miembros de la UE, al introducir la excepción al *Copyright, Designs and Patents Act* (2014), el gobierno buscaba permitir la investigación sin “socavar el control de los editores sobre los sistemas de TI o la explotación comercial” (HM Government, 2012, p. 37) [traducción libre]⁴². Según lo expresado por la Oficina de Propiedad Intelectual del Reino Unido, la creación de la excepción no significa que los investigadores tendrán acceso gratuito, ya que “los investigadores o sus instituciones todavía tienen que comprar el acceso al contenido si ese es el modelo de negocio del editor” (UK Intellectual Property Office, 2014, p. 8) [traducción libre]⁴³. No obstante, no existen restricciones sobre cómo o dónde se pueden publicar los resultados de la investigación en TDM, siempre y cuando el propósito original de la investigación no sea comercial.

No obstante, el uso no comercial no es la única restricción relevante a la excepción del Reino Unido. Existe una segunda restricción significativa respecto a la excepción de derechos de autor: el requerimiento de acceso lícito a las obras que se utilizan (Geiger *et al.*, 2018b, p. 17). El artículo 29A (1) requiere el acceso lícito para realizar la copia de la obra que se utilizará para realizar el análisis computacional. La Oficina de Propiedad Intelectual del Reino Unido define el acceso lícito como el derecho de “los investigadores para leer una obra protegida por derechos de autor” (UK Intellectual Property Office, 2014, p. 7) [traducción libre]⁴⁴. Sin embargo, como se vio anteriormente con la Directiva MUD, las limitaciones al acceso lícito no se instituyen con los contenidos de acceso abierto, sino

con los contenidos que requieren una suscripción paga. Cuando el acceso lícito está vinculado a los pagos de los servicios de suscripción, la capacidad presupuestaria del investigador llega a jugar un papel esencial, por lo que puede aumentar la brecha científica y de innovación entre los investigadores en el interior del Reino Unido.

La tercera limitación significativa proviene del artículo 29A (2), ya que la excepción solo se aplica a las reproducciones, pero no a la comunicación de la obra. A diferencia de otras excepciones de TDM que permiten la reproducción y el intercambio del trabajo, el artículo 29A (2) (a) no permite al investigador “compartir la información de investigación de TDM con ninguna otra persona” (*Copyright, Designs and Patents Act*, 1988) [traducción libre]⁴⁵, excepto cuando la transferencia esté autorizada por el propietario de los derechos de autor. La Oficina de Propiedad Intelectual del Reino Unido es clara en esta restricción, ya que las obras “no se compartirán, venderán o pondrán a disposición del público de ninguna manera y cualquiera que lo haga podría ser demandado por infracción de derechos de autor” (UK Intellectual Property Office, 2014, p. 7) [traducción libre]⁴⁶. Asimismo, aunque el artículo 5(3) de la Directiva InfoSoc (Directiva 2001/29/CE, 2001), que inspiró el artículo 29A, permite las excepciones tanto para la reproducción como para la comunicación, la ley de derechos de autor del Reino Unido decidió excluir la opción de comunicación. En un proceso de TDM, compartir información es una segunda parte importante del proceso, lo que permite que otros investigadores colaboren o incluso desarrollen nuevos proyectos de investigación a partir de la misma información. Compartir ayuda a un mayor desarrollo del conocimiento, la misma función que tiene un libro cuando se imprime y se publica para el público en general.

Como resultado, la excepción de derechos de autor de TDM del Reino Unido podría estar pendiente de revisión. Dado que la excepción se formuló antes del Brexit (2014) y se inspiró parcialmente en el artículo 5(3) de la Directiva InfoSoc (Direc-

³⁹ Véase Geiger *et al.* (2018b, p. 17) y Margoni & Dore (2016).

⁴⁰ Texto original: “a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose”.

⁴¹ Ver de forma general Rosati (2019).

⁴² Texto original: “[...] without undermining publishers’ control over IT systems or commercial exploitation”.

⁴³ Texto original: “[...] researchers or their institutions still have to buy access to content if that is the business model of the publisher”.

⁴⁴ Texto original: “researchers have the legal right to access a copyright work to read it”.

⁴⁵ Texto original: “the copy is transferred to any other person”.

⁴⁶ Texto original: “They can’t be shared, sold, or made publically available in any way and anyone doing so could be sued for copyright infringement”.

tiva 2001/29/CE, 2001), el Reino Unido podría estar *ad portas* de nuevas reglas en TDM que alineen más los enfoques internacionales actuales o, incluso, con la Directiva MUD. Sin embargo, a pesar de que la Directiva MUD de la UE podría ser un punto de partida para una reforma, incluir un uso comercial obligatorio —como se ve en Japón o Estados Unidos—, podría traer mejores resultados competitivos para los investigadores del Reino Unido. Enmarcar excepciones que pueden estar a la par con los países más progresistas también podría traer proyectos internacionales al territorio del Reino Unido.

3. Japón

Japón fue el primer país en modificar su Ley de derechos de autor en 2009 para incluir una excepción de TDM, posteriormente modificada por una enmienda de 2018 que amplió el alcance de la excepción.

La disposición TDM de 2009 se introdujo para permitir el análisis de información por computadora, autorizando la reproducción y adaptación de obras utilizadas para proyectos TDM⁴⁷. La excepción incluía una definición amplia de análisis informativo⁴⁸, no tenía restricciones de uso comercial y no era necesario que los proyectos TDM se llevarán a cabo exclusivamente con fines de investigación científica (SYNYO GmbH, 2015). Sin embargo, la excepción también contaba con algunas limitaciones relevantes: todas las tareas de TDM debían realizarse exclusivamente con un ordenador, lo que generó dudas sobre la aplicabilidad futura si entraran en juego nuevas tecnologías (Borgui & Karapapa, 2013). La excepción también parecía excluir los datos sin procesar, así como las bases de datos (Triaille *et al.*, 2014). Estas limitaciones e incertidumbres en la norma crearon la necesidad de una enmienda a la excepción de derechos de autor de TDM de 2009 (Caspers & Guibault, 2016).

La enmienda de 2018 introduce disposiciones adicionales en TDM, al tiempo que ofrece un alcance más amplio para las tecnologías involucradas⁴⁹. El artículo 30, numeral 4 (*Act 48 de 1970*, 2018), amplía el alcance de las obras para uso en TDM, al tiempo que introduce los tipos de explotación que se pueden realizar. El artículo permite la explotación de cualquier obra protegida por derechos de autor que no esté destinada a disfrutar o hacer que otra persona disfrute de las ideas o emociones expresadas en dicha obra⁵⁰. La definición de ‘disfrutar’ surge como un elemento clave para fijar el alcance de esta limitación. Para la Oficina Japonesa de Derechos de Autor (en adelante, JCO por sus siglas en inglés), el valor económico de una obra surge cuando una “[p]ersona que ve o escucha la obra, paga una compensación por dicha obra para disfrutar de las ideas o emociones expresadas en la obra, satisfaciendo los intereses intelectuales o emocionales de dicho usuario” (Japan Copyright Office, 2019, p. 12) [traducción libre]⁵¹. Por tanto, para la JCO los actos realizados sin el propósito de disfrutar de las ideas o emociones expresadas en una obra no perjudicarán los intereses del titular de los derechos de autor. El artículo 30, numeral 4, inciso (iii), proporciona ejemplos de obras no destinadas para ‘disfrute’, siendo estas obras utilizadas en el procesamiento de datos informáticos y que no involucran su percepción por medio de los sentidos humanos.

El artículo 30, numeral 4, inciso (ii), da un alcance aún más amplio a la excepción, ya que ofrece una lista completa de los pasos previsible en un proyecto de TDM. Según la Ley de Derechos de Autor japonesa, está permitido explotar el trabajo “de cualquier forma y en la medida que se considere necesario” (*Act 48 de 1970*, 2018, art. 30-4) [traducción libre]⁵². Esto incluye acciones como la “extracción, comparación, clasificación u otro análisis estadístico de datos de lenguaje, sonido o imagen, u otros elementos de los que se componga una gran cantidad de obras o un gran volumen

⁴⁷ Véase el artículo 47-7, de la Ley de Derechos de Autor de Japón (*Act 48 de 1970*, 2018):

For the purpose of information analysis [...] by using a computer, it shall be permissible to make recording on a memory, or to make adaptation (including a recording of a derivative work created by such adaptation), of a work, to the extent deemed necessary. However, an exception is made of database works which are made for the use by a person who makes an information analysis.

⁴⁸ Véase el artículo 47-7, de la Ley de Derechos de Autor de Japón (*Act 48 de 1970*, 2018):

‘[I]nformation analysis’ means to extract information, concerned with languages, sounds, images or other elements constituting such information, from many works or other such information, and to make a comparison, a classification or other statistical analysis of such information; the same shall apply hereinafter in this Article (versión del 2009).

⁴⁹ Véase de forma general Ueno (2021).

⁵⁰ Véase de forma general Ueno (2021).

⁵¹ Texto original: “[P]erson who views or listens to the work pays compensation for such work in order to enjoy the ideas or emotions expressed in a work and satisfies the person’s intellectual or emotional desires”.

⁵² Texto original: “in any way and to the extent considered necessary.”

de datos” (*Act 48 de 1970*, 2018, art. 30-4(ii)) [traducción libre]⁵³. El artículo da derecho a extraer todos los trabajos requeridos, clasificarlos y compararlos, ejecutar cualquier proceso en TDM, así como procesar análisis estadísticos de todo tipo y tamaño de trabajos, eliminando así cualquier posibilidad de restringir el procesamiento de datos.

Las excepciones de TDM de derechos de autor de Japón muestran la legislación más abierta para el uso de TDM. Contrariamente a las excepciones de la Directiva europea MUD, o el Reino Unido, Japón ofrece una excepción lo suficientemente flexible como para ofrecer un entorno abierto para la investigación y exploración de un amplio espectro de proyectos en TDM.

B. TDM basado en decisiones judiciales

1. Estados Unidos de América

Existen diversos casos que deben ser considerados para determinar la legalidad de TDM en los Estados Unidos, toda vez que esta área del derecho no está codificada en la ley. Como se verá en esta sección, casos como *Authors Guild, Inc. c. Google, Inc.* (2015) y *Authors Guild, Inc. c. Hathitrust* (2014) han construido un sólido precedente legal para permitir TDM en Estados Unidos (Sag, 2009). Esta sección se concentra en explicar el estatuto de uso justo y nombrar los casos clave para ella en este país.

a. Ley de uso justo

La doctrina del uso justo de los derechos de autor fue desarrollada inicialmente por los tribunales de los Estados Unidos para permitir la copia no autorizada de obras protegidas por derechos de autor⁵⁴. Los tribunales desarrollaron el uso justo para pro-

mover el artículo 1 de la Constitución de los Estados Unidos (en específico el inciso octavo de la octava sección) con el propósito de fomentar el “progreso de la ciencia y las artes útiles” (U.S. Const., 1791, art. 1 § 8) [traducción libre]⁵⁵. Si bien los autores son uno de los beneficiarios previstos por el derecho de autor el beneficiario primario y final es el público, cuyo acceso a las obras busca avanzar el conocimiento, mientras que se proporciona una compensación a los autores por su trabajo (U.S. Const., 1791, art. 1 § 8). Es así que los tribunales han reconocido que el objetivo final del derecho de autor es “ampliar el conocimiento y la comprensión del público”, al tiempo que otorga a los creadores el monopolio de sus obras (U.S. Const., 1791, art. 1 § 8). El uso justo de los derechos de autor surge como un desarrollo de las prioridades dadas en la Constitución de los Estados Unidos, pero también funciona como una válvula de seguridad para la Primera Enmienda (Cunard *et al.*, 2020)⁵⁶.

El uso legítimo se codificó con la sección 107 de la Ley de derechos de autor de 1976 (U.S. Code, título 17, § 107). La sección 107 formalizó el uso justo en un estatuto que incluye cuatro factores que se deben analizar a fin de determinar si existe un uso justo de una obra protegida por el derecho de autor⁵⁷. Los cuatro factores son los siguientes: (i) el propósito y el carácter del uso; (ii) la naturaleza del trabajo protegido por derechos de autor; (iii) la cantidad y sustancialidad de la parte utilizada en relación con la obra protegida por derechos de autor en su conjunto; y (iv) el efecto del uso sobre el mercado potencial o el valor de la obra protegida por derechos de autor (U.S. Code, 2012, título 17, § 107). Cada uno de los cuatro factores debe examinarse de forma independiente y luego ponderarse entre sí⁵⁸. Los cuatro factores de la sección 107 no deben “tratarse de forma aislada [...]

⁵³ Texto original: “extraction, comparison, classification, or other statistical analysis of language, sound, or image data, or other elements of which a large number of works or a large volume of data is composed.”

⁵⁴ Véase sentencia *Folsom c. Marsh* (1841) (denominado como el primer caso de uso justo); asimismo, véase de forma general Samuelson (2017).

⁵⁵ Véase el texto original: “progress of science and useful arts.”

⁵⁶ Véase también las sentencias relacionadas: *Eldred c. Ashcroft* (2003); *Harper & Row Publishers, Inc. c. Nation Enters.* (1985); y *Nat'l Rifle Ass'n of Am. c. Handgun Control Fed'n of Ohio*, (1994).

⁵⁷ Ver sección 107 del título 17 del Código de los Estados Unidos:

Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include— (1)the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2)the nature of the copyrighted work; (3)the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4)the effect of the use upon the potential market for or value of the copyrighted work (U.S. Code, 2012, título 17, § 107).

⁵⁸ El uso legítimo ha sido un concepto en constante evolución. Para el profesor Jaszi, desde la década de 1960 hasta la de 1990, el uso legítimo se había entendido como una excepción que sólo podía utilizarse “si sus actividades no invaden el mercado del titular de los derechos de autor de ninguna manera” (Aufderheide & Jaszi, 2018, p. 88). En 1990, el Juez Federal de Primera Instancia Leval cuestionó la forma en que los tribunales habían entendido —o administrado

[todos] deben explorarse, y los resultados deben sopesarse a la luz de los propósitos del derecho de autor” (*Dr. Seuss Enters., L.P. c. Penguin Books*, 1997, p. 1399) [traducción libre]⁵⁹. Cada factor es parte de la pregunta sobre

cómo definir el límite de los derechos exclusivos del autor original con el fin de servir mejor a los objetivos generales de la ley de derechos de autor de expandir el aprendizaje público mientras se protegen los incentivos de los autores para crear para el bien público⁶⁰ (*Folsom c. Marsh*, 1841) [traducción libre].

Bajo el primer factor, la corte considera hasta qué punto la nueva obra es ‘transformadora’⁶¹, ya que la obra debe agregar “algo nuevo, con otro propósito o carácter diferente, alterando el primero con nueva expresión, significado o mensaje” (*Campbell c. Acuff-Rose Music, Inc.*, 1994, p. 579) [traducción libre]⁶². Bajo el segundo factor, el tribunal considera si la obra protegida por derechos de autor es “del tipo creativo o instructivo que las leyes de derechos de autor valoran y buscan fomentar” (Leval, 1990, p. 1117) [traducción libre]⁶³. El tercer factor “pregunta si la cantidad y la sustancialidad de la parte utilizada en relación con la obra protegida

por derechos de autor en su conjunto son razonables en relación con el propósito de la copia” (*Dr. Seuss Enters., L.P. c. Penguin Books*, 1997, p. 1402) [traducción libre]⁶⁴. Finalmente, bajo el cuarto factor, los tribunales deben cuestionar

si el daño real al mercado resultó del uso del acusado [...] y si una conducta generalizada e irrestricta del tipo que lleva a cabo el acusado... resultaría en un impacto sustancialmente adverso en el potencial mercado para el original o sus derivados (*Campbell c. Acuff-Rose Music, Inc.*, 1994, p. 590) [traducción libre]⁶⁵.

En la práctica, los tribunales tienden a dar más peso a los factores uno y cuatro; pero, si el uso es de naturaleza transformadora, el primer análisis factorial debe preguntar cuál es el propósito del trabajo. Si el propósito es transformador, no debería importar si el uso es comercial o no, y se debe reconocer un uso justo⁶⁶.

C. Casos sobre uso justo y minería de textos y datos (TDM)

Actualmente, dos fallos judiciales en Estados Unidos permiten el uso de TDM: el caso *Authors*

intuitivamente— el uso justo, y abogó por una interpretación más centrada en el propósito transformador que en el efecto del uso en el mercado potencial o el valor del producto o trabajo protegido por derechos de autor (“Judges do not share a consensus on the meaning of fair use [...]. The opinions reflect widely differing notions [...] [d]ecisions are not governed by consistent principles, but seem to result from intuitive reactions to individual fact patterns”) (Leval, 1990, pp. 1106-1107). Asimismo, ver el caso *Campbell c. Acuff-Rose Music, Inc.*, (1994), en el que se sostiene que un propósito transformador hace que el primer factor sea más probable que favorezca el uso justo. Desde *Campbell*, la lectura del uso legítimo se ha ampliado para dar más peso al elemento transformador—el primer factor—, enfoque comúnmente seguido por la Corte de Apelaciones del Segundo Circuito de los Estados Unidos.

⁵⁹ Texto original: “The four fair use factors ‘are to be weighed together, in light of the objectives of copyright to promote the progress of science and the useful arts’”.

⁶⁰ Véase la sentencia *Folsom c. Marsh* (1841) (denominado como el primer caso de uso justo); asimismo, véase de forma general Samuelson (2017).

⁶¹ Véase *Mattel, Inc. c. Walking Mountain Prods.* (2003).

⁶² Texto original: “something new, with a further purpose or different character, altering the first with new expression, meaning, or message”.

⁶³ Texto original: “consider whether the protected writing is of the creative or instructive type that the copyright laws value and seek to foster”. Este factor rara vez juega un papel importante en la determinación de una disputa por uso justo. Véase también Patry (2015).

⁶⁴ Texto original: “whether the amount and substantiality of the portion used in relation to the copyrighted work as a whole was reasonable to the purpose of the copying”.

⁶⁵ Texto original:

market harm caused by the particular actions of the alleged infringer, but also “whether unrestricted and widespread conduct of the sort engaged in by the defendant... would result in a substantially adverse impact on the potential market” for the original.

⁶⁶ Con el caso *Campbell c. Acuff-Rose Music, Inc.* (1994) los tribunales comenzaron a dar más peso al elemento transformador en sus consideraciones de uso justo, véase *Authors Guild, Inc. c. Google, Inc.* (2015, p. 15):

Notwithstanding fair use’s long common-law history, not until the Campbell ruling in 1994 did courts undertake to explain the standards for finding fair use [...]. The Campbell Court undertook a comprehensive analysis of fair use’s requirements, discussing every segment of § 107 [...]. [T]he Court made clear that they are “illustrative and not limitative” and “provide only general guidance about the sorts of copying that courts and Congress most commonly ha[ve] found to be fair uses.” [...]. The statute “calls for case-by-case analysis” and “is not to be simplified with bright-line rules.

Ver también *Castle Rock Entm’t, Inc. c. Carol Pub. Grp., Inc.* (1998) y *Cariou c. Prince* (2013). Asimismo, ver Samberg & Hennesy (2018).

Guild, Inc. c. HathiTrust (2014) y *Authors Guild, Inc. c. Google* (2015)⁶⁷. En este último, el Tribunal permitió que realizara una digitalización no autorizada de obras protegidas por derechos de autor, la creación de una función de búsqueda y la visualización de fragmentos de esas obras en los motores de búsqueda de Google (*Authors Guild, Inc. c. Google*, 2015). En el *Authors Guild, Inc. c. HathiTrust* (2014), la Corte permitió que las bibliotecas del proyecto HathiTrust digitalizarán obras protegidas por derechos de autor con el fin de permitir búsquedas de texto completo. También autorizó a las bibliotecas miembros del proyecto proporcionar a los usuarios con discapacidades el acceso a las obras de texto completo, y que las bibliotecas conservarán los libros protegidos por derechos de autor en formato digital.

Ambos casos fueron decididos en apelación por el Tribunal Federal de Apelaciones del Segundo Circuito y los usos se consideraron lo suficientemente transformadores como para estar cubiertos por la excepción de uso justo de los derechos de autor. En *Authors Guild, Inc. c. HathiTrust*, la Corte concluyó que la creación de una base de datos de búsqueda de texto completo es un uso transformador: cuando se realiza una búsqueda de palabras, el resultado es diferente en “propósito, carácter, expresión, significado y mensaje de la página (y el libro) de la que se extrae” (2014, p. 18) [traducción libre]⁶⁸. Por lo tanto, hay poca o ninguna semejanza entre el trabajo original y los resultados de la búsqueda de texto completo de la Biblioteca digital HathiTrust.

En el caso Google, el Tribunal también determinó que la creación por parte de la empresa de una copia digital de los libros de los demandantes con fines de búsqueda es transformadora: “[e]l propósito de permitir una búsqueda para la identificación de libros que contienen un término de interés para el buscador implica un propósito altamente transformador” (*Authors Guild, Inc. c. Google, Inc.*, 2015, p. 21) [traducción libre]⁶⁹. El propósito de Google era poner a disposición información significativa sobre los libros de obras copiadas, lo que permitía a los buscadores encontrar las obras que contenían una palabra de interés; además, la función ‘ngrams’ permitió a los lectores conocer la frecuencia de uso en el *corpus* agregado de libros publicados para diferentes períodos históricos. Esas dos funciones demostraron ser lo suficientemente transformadoras como para favorecer el uso justo de las obras protegidas por derecho de autor.

Es importante señalar que los casos de HathiTrust y Google muestran cómo se permite el uso comercial y no comercial de TDM en los Estados Unidos: HathiTrust era una entidad educativa sin fines de lucro, mientras que Google es una corporación comercial con fines de lucro⁷⁰. En ambos casos, se permitió la minería de textos y datos (TDM) después de que se cumplieron los requisitos de transformación de uso justo. En este sentido, si el nuevo trabajo agrega algo nuevo con un propósito adicional o carácter diferente –alterando el primero con una nueva expresión, significado o mensaje–, o si el trabajo cumple con los requisitos transformado-

⁶⁷ Ver de forma general a Carroll (2019); ver también el caso *Williams & Wilkins Co. c. Estados Unidos* (1973). Este es el primer caso relevante sobre copias no autorizadas institucionales y sistemáticas con fines de investigación. En el caso *Williams & Wilkins Co. c. Estados Unidos*, un editor de revistas médicas demandó a un instituto de investigación médica del gobierno y a su biblioteca por infracción de derechos de autor. El Departamento de Salud, Educación y Bienestar, a través de los Institutos Nacionales de Salud (NIH) y la Biblioteca Nacional de Medicina (NLM), hizo fotocopias no autorizadas de algunos de los artículos de las revistas científicas de la editorial Medicine, Journal of Immunology, Gastroenterology y Revisiones farmacológicas. Así pues, los investigadores médicos que solicitaron las fotocopias eran investigadores científicos y profesionales que necesitaban los artículos para uso personal en su trabajo científico, y no había intención de duplicarlos para la venta u otra distribución general. Las instituciones también restringieron la copia única de un solo artículo con una extensión limitada de cincuenta páginas. Sobre la base del tipo y contexto de uso, el Tribunal no encontró ninguna infracción, ya que el uso impugnado se consideró como uso legítimo. Así, el razonamiento del tribunal utilizó principalmente el cuarto factor del uso legítimo para llegar a su decisión. El demandante no pudo probar su presunción de perjuicio económico, en el pasado o potencialmente para el futuro. Como señaló el Tribunal, “[the] record simply does not show a serious adverse impact, either on plaintiff or on medical publishers generally, from the photocopying practices of the type of NIH and NLM” (*Williams & Wilkins Co. c. Estados Unidos*, 1973, p. 1359). Por lo tanto, el editor no demostró como un hecho probado la hipótesis hipotética de pérdida de mercado. Pero el Tribunal consideró que se podría causar un perjuicio a la investigación médica y científica si se consideraba ilegal la fotocopia de este tipo. Así, el Tribunal falló a favor de los acusados. Por todo esto, este caso cobró relevancia, ya que se autorizó a las bibliotecas a realizar copias con fines de investigación.

Véanse también otros casos de copia sistemática e institucional de imágenes con fines transformadores cuando las copias cumplían una función diferente a la original: *Kelly c. Arriba Soft Corp.* (2003); *Perfect 10, Inc. c. Amazon.com* (2007); *A.V. ex rel. Vanderhye c. iParadigms, LLC* (2009).

⁶⁸ Texto original: “purpose, character, expression, meaning, and message from the page (and the book) from which it is drawn”.

⁶⁹ Texto original: “[the] purpose of enabling a search for identification of books containing a term of interest to the searcher involves a highly transformative purpose”.

⁷⁰ Véase *Authors Guild, Inc. c. Google, Inc.* (2015) y *Authors Guild, Inc. c. HathiTrust* (2014).

res del primer factor de uso justo, menor será la importancia de otros factores a fin de determinar si existe un uso legítimo⁷¹.

Aunque los casos actuales permiten el uso comercial y no comercial de TDM, aún quedan preguntas sin resolver (Sag, 2019). No existe una respuesta específica sobre cómo abordar los problemas que surgen en virtud del derecho contractual, las leyes que prohíben la elusión de las medidas de protección tecnológica o los problemas transfronterizos de derechos de autor ¿Qué sucede cuando se desarrolla un proyecto de investigación TDM con catálogos de bibliotecas en diferentes países? ¿Pueden los investigadores desarrollar proyectos TDM sin acceso legal a los datos o la información? ¿Es posible participar en proyectos internacionales y compartir los resultados de los contenidos? ¿Pueden los bibliotecarios negociar para eliminar o reducir las restricciones contractuales sobre los derechos de los investigadores para participar en TDM? A medida que avanza la tecnología, seguirán surgiendo más preguntas sobre cómo nivelar las leyes de derechos de autor con el acceso de los usuarios y la innovación.

IV. FUTURO DEL DESARROLLO DE POLÍTICAS LEGISLATIVAS EN MINERÍA DE TEXTOS Y DATOS

Una legislación integral en TDM se presenta como una herramienta esencial para los investigadores, ya que puede permitir el acceso legal a materiales y bases de datos, así como el derecho de reproducción, almacenamiento y comunicación al público sin que se presenten posibles vulneraciones a los derechos de autor (Francoise, 2020). Una excepción en TDM adecuadamente estructurada puede permitir a los investigadores desarrollar proyectos sobre esta con materiales de acceso abierto y contenidos basados en suscripción; puede permitir TDM para uso comercial y no comercial; y construir análisis y proyectos de TDM no limitados a propósitos de investigación científica⁷². Asimismo, una excepción abierta al respecto debería permitir el almacenamiento del *corpus* o las copias generadas con TDM, además de tener la opción de compartir los resultados del proyecto con terceras personas y sin limitarlo a un círculo privado de contactos. En general, una excepción en TDM puede brindar acceso abierto a todo tipo de usuarios, incluidos, entre otros, investigadores, académicos, organizaciones sin fines de lucro o con fines de lucro.

No obstante, como se vio en las secciones anteriores, las leyes regionales o nacionales aprobadas en TDM carecen de muchas de las características descritas anteriormente. Por ejemplo, los artículos de TDM aprobados en la Directiva europea del mercado único digital (Directiva 2019/790, 2019) serán incompatibles en algunas partes con la reforma de derechos de autor de TDM japonesa de 2018 o las recientes sentencias de Google o HathiTrust del Tribunal de Apelaciones del Segundo Circuito de los Estados Unidos. Mientras que Europa y el Reino Unido distinguen el uso comercial y no comercial, Japón y Estados Unidos no. Además, mientras que Alemania limita el intercambio a un grupo cercano de investigadores, Japón permite el intercambio gratuito sin restricciones. Si bien el uso comercial en Europa puede excluirse mediante contratos, en el Reino Unido, el uso comercial ni siquiera forma parte de las excepciones de TDM.

Asimismo, la mayoría de los países en desarrollo no han iniciado el proceso de actualización de sus leyes de derecho de autor para incluir excepciones que permitan la TDM. La única alternativa con la que cuentan los países sin una legislación dedicada sobre este punto es la adopción de interpretaciones flexibles y abiertas sobre las L&E existentes como el derecho a la investigación, la reprografía digital o el uso privado o personal. Colombia posiblemente podría contar con flexibilidades en su legislación para el uso privado (Ley 23, 1982, art. 37), pero otras legislaciones, como la de Perú (Decreto Legislativo 822, 1996, art. 43 & 48) o Nicaragua (Ley 312, 1999, art. 31), contarían con restricciones sobre la cantidad de uso que se puede hacer sobre una obra. Igualmente, no tener reglas explícitas al respecto incrementa los niveles de incertidumbre dados los nuevos y constantes desarrollos tecnológicos de IA para explorar y explotar la información. Además, los países con una interpretación restrictiva de L&E, como Argentina (Ley 11.723, 1933, art. 10), sin duda prohibirán a los investigadores permitir cualquier TDM sobre obras protegidas por el derecho de autor.

Como lo menciona Okediji (2017), existen beneficios relevantes para la armonización de las leyes básicas de derechos de autor, en este caso, para permitir el flujo transfronterizo de información. La TDM requiere un entorno internacional integral de derechos legales para crecer como una fuente mundial de conocimiento. Al revisar las posibles disposiciones legales de derechos de autor invo-

⁷¹ Ver *Campbell c. Acuff-Rose Music, Inc.* (1994) y también *Authors Guild, Inc. c. Google, Inc.* (2015).

⁷² Véase Flynn *et al.*: "TDM need not come at the expense of right holders. These reproductions do not compromise the core interests of exclusive rights, which is to prohibit unauthorised reproductions that can substitute for the work of the author" (2020, p. 4).

lucradas en el proceso de aprendizaje automático, las siguientes consideraciones están disponibles.

En primer lugar, una política legislativa debe cuestionar si el acceso lícito debe ser un requisito para la TDM. El acceso legal cubre el contenido basado en una política de acceso abierto, o mediante acuerdos contractuales entre los titulares de derechos y los beneficiarios de la excepción. La política de acceso abierto se refiere “a la información científica que es gratuita para el usuario y que es reutilizable” (European Commission, 2020b) [traducción libre]⁷³, mientras que el contenido del acuerdo contractual cubre todas las suscripciones y el contenido que está disponible gratuitamente en línea. No obstante, las limitaciones al acceso legal no vienen con los contenidos de acceso abierto, sino con los contenidos que requieren una suscripción paga. Cuando el acceso legal está vinculado a los pagos de los servicios de suscripción, la capacidad presupuestaria de la organización pasa a jugar un papel fundamental. Un posible resultado es que solo unas pocas organizaciones de investigación podrán adquirir licencias para todas las bases de datos que son relevantes para un proyecto de investigación en TDM, lo que a cambio limitará el desarrollo de TDM para un número relevante de organizaciones de investigación. Si la capacidad de financiamiento de cada institución determina los resultados de la investigación en TDM, el acceso lícito puede aumentar las diferencias entre las instituciones, poniendo a muchas de ellas en condiciones menos favorables.

En segundo lugar, una política legislativa debe cuestionar si es necesario eliminar la división entre uso comercial y no comercial. Si bien la UE divide el régimen TDM entre uso no comercial y comercial, otros países como Estados Unidos y Japón han evitado ese enfoque, permitiendo TDM sin importar el propósito comercial o el tipo de usuario. Los fallos judiciales de TDM de los Estados Unidos, *Authors Guild, Inc. c. Google, Inc.* (2015) y *Authors Guild, Inc. c. HathiTrust* (2014), han permitido el uso comercial de la investigación de TDM bajo excepciones de uso justo. Japón no distingue entre uso comercial y no comercial. El análisis de la información no tiene que realizarse necesariamente con fines de investigación científica y la TDM está abierta a todos los usuarios. Este enfoque sigue el objetivo de Japón de promover el desarrollo de las industrias de IA y el *big data*. En consecuencia, legislaciones que restrinjan la TDM en sus territorios al uso no comercial, eventualmente podrían ver una fuga de proyectos comerciales a otros países donde ese tipo de uso sí esté permitido.

En tercer lugar, una política legislativa debe incluir el almacenamiento, la comunicación y distribución de datos. En un proyecto TDM, compartir información es una segunda parte importante del proceso, ya que permite que otros investigadores colaboren o incluso desarrollen nuevos proyectos de investigación a partir de la misma información. Compartir ayuda a un mayor desarrollo del conocimiento, la misma función que tiene un libro cuando se imprime.

V. CONCLUSIONES

El panorama ofrece un escenario muy desigual, dividido e incierto para los investigadores de todo el mundo. Como cada investigador depende de sus leyes de derechos de autor nacionales o regionales, los resultados del acceso a los materiales siguen siendo restringidos. De no contar con instrumentos armoniosos, estos colocarán a los investigadores y usuarios en posiciones desfavorables, ya que la investigación de TDM en diferentes campos del conocimiento seguirá dependiendo de las restricciones de las variables de derechos de autor. Además, la falta de una legislación uniforme podría aumentar la brecha científica y de innovación entre las naciones desarrolladas y las menos desarrolladas.

Japón y Estados Unidos son un ejemplo interesante sobre cómo legislar en TDM. No obstante, aun así, para que la TDM fomente la innovación, los países seguramente tendrán que eliminar barreras como el acceso lícito, la división entre uso comercial y no comercial, la limitación para compartir los datos, el *corpus*, la investigación o los resultados de los proyectos de TDM, así como cualquier limitación para compartir y comunicar la información recogida y obtenida a partir de los proyectos en TDM.

La mayoría de los países en desarrollo no han iniciado el proceso de actualización de sus leyes de derecho de autor para incluir excepciones que permitan la TDM. La única alternativa con la que cuentan los países sin una legislación dedicada sobre este punto es la adopción de interpretaciones flexibles y abiertas sobre las L&E existentes, como el derecho a la investigación, la reprografía digital o el uso privado o personal.

Asimismo, no tener reglas explícitas al respecto incrementa los niveles de incertidumbre dados los nuevos y constantes desarrollos tecnológicos de IA para explorar y explotar la información. Los países con una interpretación restrictiva en sus excepciones y limitaciones, sin duda alguna, prohibirán a los investigadores permitir cualquier TDM sobre obras protegidas por el derecho de autor. 🗑️

⁷³ Texto original: “scientific information that is free of charge to the user and is reusable”.

REFERENCIAS

- Aufderheide, P., & Jaszi, P. (2018). *Reclaiming Fair Use* (2da ed.). University of Chicago Press.
- Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, 1(3), 295-311. <https://doi.org/10.1162/neco.1989.1.3.295>
- Borghi, M. (2016). *Text and Data Mining*. <https://www.copyrightuser.org/understand/exceptions/text-data-mining/>
- Borghi, M., & Karapapa, S. (2013). *Copyright and Mass Digitization: A Cross-Jurisdictional Perspective*. Oxford University Press.
- Carroll, M. (2019). Copyright and the Progress of Science: Why Text and Data Mining Is Lawful. *UC Davis Law Review*, 53(2), 893-964.
- Caspers, M., Guibault, L., McNeice, K., Piperidis, S., Pouli, K., Eskevich, M., & Gavrilidou, M. (2017). *Baseline report of policies and barriers of TDM in Europe (extended version)*. FutureTDM project report. https://www.futuretdm.eu/wp-content/uploads/FutureTDM_D3.3-Baseline-Report-of-Policies-and-Barriers-of-TDM-in-Europe.pdf
- Caspers, M., & Guibault, L. (2016). *Deliverable 3.3+ Baseline report of policies and barriers of TDM in Europe*. https://project.futuretdm.eu/wp-content/uploads/2017/05/FutureTDM_D3.3-Baseline-Report-of-Policies-and-Barriers-of-TDM-in-Europe.pdf
- Castaneda, R., Arroyo, P., & Loza, L. (2020). Assessing Countries Sustainability: A Group Multicriteria Decision Making Methodology Approach. *Journal of Management and Sustainability*, 10(1), 174-188.
- Castle, N. (13 de julio de 2017). Supervised vs. Unsupervised Machine Learning [Entrada de blog]. <https://www.datascience.com/blog/supervised-and-unsupervised-machine-learning-algorithms>
- Coglianesi, C., & Lehr, D. (2017). Regulating by Robot: Administrative Decision Making in the Machine-Learning Era. *The Georgetown Law Journal*, 105, 1147-1223.
- Congressional Research Service [CRS] (2017). *Laurie A. Harris, Overview of Artificial Intelligence*. <https://crsreports.congress.gov/product/pdf/IF/IF10608/3>.
- Cunard, J. P., Keller, B., & Potenza, M. (2020). *Copyright Fair Use. Copyright*. Thomsons Reuters Practical Law.
- Davis, J. (14 de octubre de 2020). Zillow uses Analytics, Machine Learning to Disrupt with Data. *Information Week*. <http://www.informationweek.com/big-data/zillow-uses-analytics-machine-learning-to-disrupt-with-data/d/d-id/1327175>
- Ducato, R., & Strowel, A. (19 de marzo de 2019). Limitations to Text and Data Mining and Consumer Empowerment. Making the Case for a Right to 'Machine Legibility'. [Entrada de blog]. <http://copyrightblog.kluweriplaw.com/2019/03/19/limitations-to-text-and-data-mining-and-consumer-empowerment-making-the-case-for-a-right-to-machine-legibility/>
- Dusollier, S. (2003). Tipping the Scale in Favor of the Right Holders: The European Anti-Circumvention Provisions. En E. Becker, W. Bushé, D. Günnewig & N. Rump (eds.), *Digital Rights Management* (pp. 462-478). Springer.
- Eskevich, M., & Van Den Bosch, A. (2016). *Deliverable D3.1: Research Report on TDM Landscape in Europe*. <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5a5c961d0&appId=PPGMS>
- European Parliamentary Research Service [EPRS] (2020). *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf)
- European Commission (2020a). *Shaping Europe's digital future*. <https://ec.europa.eu/digital-single-market/en/copyright>
- (2020b). *Open Access*. https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/open-access_en
- Fagan, F. (2015). From Policy Confusion to Doctrinal Clarity: Successor Liability from the Perspective of Big Data. *Virginia Law & Business Review*, 9(3), 391-451.
- Flynn, S., Geiger, C., Quintais, J., Margoni, T., Sag, M., Guibault, L., & Carroll, M. (2020). Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for In-

- ternational Action. *European Intellectual Property Review*, (7), 1-12. <https://dx.doi.org/10.2139/ssrn.3578819>
- Flynn, S., Izquierdo, H.A., & Palmedo, M. (2021). *Research for Text and Data Mining in Comparative Copyright Law*. Program on Information Justice and Intellectual Property, Forthcoming.
- Francoise, C. (10 de Junio de 2020). Text and Data Mining: (Articles 3 and 4 of the EU-DSM) by REBIUN's Copyright working group [Entrada de blog]. <https://blogs.ifla.org/lpa/2020/06/10/text-and-data-mining-articles-3-and-4-of-the-eu-dsm/>
- Geiger, C. Frosio, G., & Bulayenko, O. (2018a). The EU Commission's Proposal to Reform Copyright Limitations: A Good but Far Too Timid Step in the Right Direction. *European Intellectual Property Review*, 40(1), 4-15.
- (2018b). Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?. *International Review Intellectual Property and Competition Law*, 49(7), 814-844.
- (2018c). *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects* (Centre for International Intellectual Property Studies (CEIPI) Research Paper No. 2018-02). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3160586.
- (2019). Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU. En C. Saiz García & R. Evangelio Llorca (eds.), *Propiedad Intelectual y Mercado Único Europeo* (pp. 23-64). Tirant lo Blanch.
- Guerra, P., & Tamburello, P. (2018). *Modernizing Cybersecurity Operations with Machine Intelligence, Advanced Threat Detection, Hunting, and Analysis*. O' Reilly Media Inc. <https://www.oreilly.com/library/view/modernizing-cybersecurity-operations/9781492035992/>
- Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M., (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2), 245-258. <https://doi.org/10.1016/j.neuron.2017.06.011>
- Hilty, R., Harhoff, D., & Drexler, J. (2016). *Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules*. https://www.ip.mpg.de/fileadmin/ipmpg/content/stellungnahmen/MPI_Position_Statement_Part_B_Chapter_1_Update23022017.pdf
- Hinton, G. E., & Sejnowski, T. J. (eds.) (1999). *Unsupervised learning: Foundations of neural computation in Unsupervised learning: Foundation computation*. MIT Press.
- HM Government (2012). *Modernising Copyright: A modern, robust and flexible framework, Government response to consultation on copyright exceptions and clarifying copyright law*. https://www.mpaonline.org.uk/wp-content/uploads/2017/09/Modernising_Copyright_-_a_modern_robust_and_flexible_framework_-_Government_response.pdf.
- Hughenoltz, B. (24 de julio de 2019). The New Copyright Directive: Text and Data Mining (Articles 3 and 4) [Entrada de blog]. <http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/>
- Japan Copyright Office [JCO] (2019). *Outline of the Amendments to the Copyright Act in 2018*.
- Le, Q. V., Ranzato, M., Monga, R., Devin, M., Chen, K., Corrado, G., Dean, J., & Ng, A.Y. (2012). *Building High-level Features Using Large Scale Unsupervised Learning*. Trabajo presentado en 29th International Conference on Machine Learning, Edimburgo, Escocia. https://static.googleusercontent.com/media/research.google.com/en//archive/unsupervised_icml2012.pdf
- Lehr, D., & Ohm, P. (2017). Playing with the Data: What Legal Scholars Should Learn About Machine Learning. *UC Davis Law Review*, 51(2), 653-717.
- Leval, P. N. (1990). Toward a Fair Use Standard. *Harvard Law Review*, 103(5), 1105-1136.
- Litman, J. (1994). The Exclusive Right to Read. *Cardozo Arts & Entertainment Law Journal*, 13(29), 34-43.
- Margoni T., & Dore, G. (2016). *Why We Need a Text and Data Mining Exception (But it is Not Enough)*. Trabajo presentado en Language Resources and Evaluation Conference (LREC 2016, 10th ed.), Portorož, Slovenia. <https://doi.org/10.5281/zenodo.248048>

- Metz, C. (4 de setiembre de 2020). Now Anyone Can Tap the AI Behind Amazon's Recommendations. *WIRED*. <http://www.wired.com/2015/04/now-anyone-can-tap-ai-behind-amazons-recommendations/>
- MIT Lincoln Laboratory [MIT LL] (2019). *Artificial Intelligence: Short History, Present Developments, and Future Outlook*. <https://www.ll.mit.edu/r-d/publications/artificial-intelligence-short-history-present-developments-and-future-outlook-0>
- MIT Sloan Management Review [MIT SMR] (2020). *Connections Custom Research Report, How AI Changes the Rules: New Imperatives for the Intelligent Organization, SAS*. <https://sloanreview.mit.edu/sponsors-content/how-ai-changes-the-rules-new-imperatives-for-the-intelligent-organization/>
- Mokhtarian, E. (2018). The Bot Legal Code: Developing a Legally Compliant Artificial Intelligence. *Vanderbilt Journal of Entertainment and Technology Law*, 21(1), 145-207.
- Myers, S. (27 de abril de 2007). Estonia removes Soviet-era war memorial after a night of violence. *New York Times*. <https://www.nytimes.com/2007/04/27/world/europe/27iht-estonia.4.5477141.html>
- New America (2017). *Full Translation: China State Council, A Next Generation Artificial Intelligence Development Plan*. <https://www.newamerica.org/documents/1959/translation-fulltext-8.1.17.pdf>
- Office of Scholarly Communication (2013). *UC Systemwide Academic Senate Open Access Policy*.
- Okediji, R. (ed.) (2017). *Copyright Law in an Age of Exceptions and Limitations*. Cambridge University Press.
- Organización Mundial de Propiedad Intelectual (2020). *WIPO Conversation on Intellectual Property and Artificial Intelligence*. https://www.wipo.int/about-ip/en/artificial_intelligence/policy.html#:~:text=The%20WIPO%20Conversation%20on%20IP,that%20policymakers%20need%20to%20ask
- PubMed Central (PMC) (s/f). *US National Library of Medicine National Institutes of Health*. <https://www.ncbi.nlm.nih.gov/pmc/about/intro/>
- Quintais, J. (2020). The New Copyright in the Digital Single Market Directive: A Critical Look. *European Intellectual Property Review*, (1), 1-23. <https://dx.doi.org/10.2139/ssrn.3424770>
- Rosati, E. (2019). *Copyright and the Court of Justice of the European Union*. Oxford University Press.
- Sag, M. (2009). Copyright and Copy-Reliant Technology. *Northwestern University Law Review*, 103(4), 1607-1682.
- (2019). The New Legal Landscape for Text Mining and Machine Learning. *Journal of the Copyright Society of the USA*, 66, 1-64.
- Samberg, R., & Hennesy, C. (2018). Law & Literacy In Non-consumptive Text Mining: Guiding Researchers Through The Landscape Of Computational Text Analysis. En S. Benson (ed.), *Copyright Conversation: Rights Literacy, in A Digital World* (pp. 289-315). UC Berkeley Previously Published Works.
- Samuelson, P. (2017). Justifications for Copyright Limitations & Exceptions. En R. Okediji (ed.), *Copyright Law in an Age of Limitations and Exceptions*, 12-60. Nueva York: Cambridge University Press. <https://doi.org/10.1017/9781316450901.003>
- (12 de julio de 2018). The EU's Controversial Digital Single Market Directive – Part II: Why the Proposed Mandatory Text- and Data-Mining Exception Is Too Restrictive [Entrada de blog]. <http://copyrightblog.kluweriplaw.com/2018/07/12/eus->
- Select Committee on Artificial Intelligence of the National Science & Technology Council (2019). *The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update*. <https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf>
- Spielkamp, M. (2017). Inspecting algorithms for bias. *MIT Technology Review*, 120(4), 96-98.
- SYNYO GmbH, (2015). *Reducing Barriers and Increasing Uptake of Text and Data Mining for Research Environments using a Collaborative Knowledge and Open Information Approach*. https://www.futuretdm.eu/wp-content/uploads/FutureTDM_D3.3-Baseline-Report-of-Policies-and-Barriers-of-TDM-in-Europe-1.pdf
- Triaille, J., Meeûs d'Argenteuil, J., & Francquen, A. (2014). *A Study on the legal framework*

of text and data mining (TDM). <https://op.europa.eu/en/publication-detail/-/publication/074ddf78-01e9-4a1d-9895-65290705e2a5/language-en>

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433-460.

U.S. Senate, Subcommittee On Space, Science, And Competitiveness, Committee On Commerce, Science, And Transportation (2016). *The Dawn of Artificial Intelligence: Hearing Before the Subcomm. on Space, Sci., & Transp.* (114th Congress) (pp. 114-562). <https://www.govinfo.gov/content/pkg/CHRG-114shrg24175/pdf/CHRG-114shrg24175.pdf>

Ueno, T. (2019). TDM Exception in Japan, Possible implication for Europe?. Trabajo presentado en Challenges of the Directive 2019/790 on Copyright and Related Rights in the Digital Single Market de la Facultad de Derecho de la Universidad Jagellónica, Cracovia, Polonia. https://ipwi.uj.edu.pl/documents/122195199/144296432/Ueno_2019_Cracow_Comment.pdf/adf-57b6a-ad86-4bd6-94b0-4bbac61b049c

(2021). The Flexible Copyright Exception for “Non-Enjoyment” Purposes- Recent Amendment in Japan and its Implication. *GRUR International*, 70(2), 145-152.

UK Intellectual Property Office (2014). *Exceptions to copyright: Research*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf

Zhongzhi, S. (2011). Advanced Artificial Intelligence. *Series on Intelligence Science*, 1. <https://ishare.iask.sina.com.cn/f/34881213.html>

LEGISLACIÓN, JURISPRUDENCIA Y OTROS DOCUMENTOS LEGALES

A.V. ex rel. Vanderhyye c. iParadigms, LLC, 562 F.3d 630, 640 (4th Cir. 2009).

Authors Guild, Inc. c. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

Authors Guild, Inc. c. HathiTrust, 755 F.3d 87 (2d Cir. 2014).

Campbell c. Acuff-Rose Music, Inc., 510 U.S. 569 (1994).

Cariou c. Prince, 714 F.3d 694, 708 (2nd Cir. 2013).

Castle Rock Entm't, Inc. c. Carol Pub. Grp., Inc., 150 F.3d 132, 141-42 (2d Cir.1998).

Convenio de Berna para la Protección de las Obras Literarias y Artísticas, 9 de septiembre de 1886, revisado en París 24 de julio de 1971, 1161 U.N.T.S. 3.

Copyright Act of 1976, 17 U.S. Code § 107 (2012) (Limitaciones a los derechos exclusivos: Uso justo) (EE.UU.).

Copyright, Designs and Patents Act 1988, c. 48 (Reino Unido).

Decreto Legislativo 822, Ley sobre el Derecho de Autor, 23 de abril de 1996 (Perú).

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, 1996 O.J. (L 77) 20 (UE).

Directiva 2001/29/CE del Parlamento Europeo y del Consejo, de 22 de mayo de 2001, relativa a la armonización de determinados aspectos de los derechos de autor y derechos afines a los derechos de autor en la sociedad de la información, 2001 O.J. (L 160) 10 (UE).

Directiva 2017/1564 del Parlamento Europeo y del Consejo, de 13 de setiembre de 2017, sobre ciertos usos permitidos de determinadas obras y otras prestaciones protegidas por derechos de autor y derechos afines en favor de personas ciegas, con discapacidad visual o con otras dificultades para acceder a textos impresos, y por la que se modifica la Directiva 2001/29/CE relativa a la armonización de determinados aspectos de los derechos de autor y derechos afines a los derechos de autor en la sociedad de la información, 2017 O.J. (L 242) 6 (UE).

Directiva 2019/790 del Parlamento Europeo y del Consejo, de 17 de mayo 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 (Directiva del mercado único digital), 2019 O.J. (L 130) 92 (UE).

Dr. Seuss Enters., L.P. c. Penguin Books, 109 F.3d 1394, 1399 (9th Cir.1997).

Eldred c. Ashcroft, 537 U.S. 186, 219-20 (2003).

- Folsom c. Marsh, 9. F. Cas. 342 (C.C.D. Mass. 1841).
- Harper & Row Publishers, Inc. c. Nation Enters., 471 U.S. 539, 560 (1985).
- Kelly c. Arriba Soft Corp., 336 F.3d 811, 822 (9th Cir. 2003).
- L. 23/1982, 19 de febrero de 1982, Diario Oficial [D.O.] 35.949 (Colom.).
- Ley 11.723, Régimen Legal de la Propiedad Intelectual [RLPI], 30 de septiembre de 1933, B. O. 11793 (Arg.)
- Ley 312, 6 de julio de 1999, Ley de Derechos de Autor y Derechos Conexos [LDADC], La Gaceta, Diario Oficial [L.G.], 1 de setiembre de 1999 (Nicar.).
- Ley 4 de 1998, Ley sobre Derechos de Autor (Cap. 300, modificada por la Ley 1 de 2006) (Barb.).
- Ley de Derecho de Autor, Act 48 de 6 de mayo de 1970 (modificado por Act 72 de 13 de julio de 2018) (Japón).
- Ley de Derecho de Autor [LDA], Capítulo 252 de las Leyes de Belice, 31 de diciembre de 2000 (Belice).
- Mattel, Inc. c. Walking Mountain Prods., 353 F.3d 792, 800 (9th Cir.2003).
- Nat'l Rifle Ass'n of Am. c. Handgun Control Fed'n of Ohio, 15 F.3d 559, 562 (6th Cir. 1994).
- Perfect 10, Inc. c. Amazon.com, Inc., 508 F.3d 1146, 1169 (9th Cir. 2007).
- Rb Dan Haag 5 februari 2020, NJ 2020, 368 m.nt. van E.J. Dommering (NJCM et al./Países Bajos) (Países Bajos).
- Reglamento 2017/1563 del Parlamento Europeo y del Consejo, de 20 de setiembre de 2017, sobre el intercambio transfronterizo entre la Unión y terceros países de ejemplares en formato accesible de determinadas obras y otras prestaciones protegidas por derechos de autor y derechos afines en favor de personas ciegas, con discapacidad visual o con otras dificultades para acceder a textos impresos, 2017, O.J. (L 242) 1 (UE).
- Urheberrechtsgesetz [UrhG] [Ley de Derechos de Autor y Derechos Conexos], 9 de setiembre de 1965, BGBl I en 1273, modificado por el artículo 1 de la Ley de 1 de setiembre de 2017 (Alemania).
- Williams & Wilkins Co. c. United States, 487 F.2d 1345 (Ct. Cl. 1973), aff'd by an equally divided Court, 420 U.S. 376 (1975).