

¿PUEDEN LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL «APRENDER» UTILIZANDO OBRAS PROTEGIDAS POR DERECHOS DE PROPIEDAD INTELECTUAL SIN LA AUTORIZACIÓN DE SUS TITULARES? LOS RETOS DEL *MACHINE LEARNING* EN PERSPECTIVA COMPARADA: UE, EEUU Y JAPÓN

Por Marta NADAL CEBRIÁN
Abogada de Uría Menéndez
LL.M. Master in Legal Tech
IE Law School

Fecha de recepción: 17.08.2021
Fecha de aceptación: 19.09.2021

RESUMEN:

El «aprendizaje» por parte de los sistemas de inteligencia artificial (particularmente, de machine learning) plantea un reto importante para la propiedad intelectual. La razón de ello es que, en buena parte de las ocasiones, estos sistemas son entrenados con obras protegidas por derechos de propiedad intelectual. En la medida en que en ese proceso de entrenamiento tienen lugar varias actividades que afectan a los derechos exclusivos de los titulares, principalmente el derecho de reproducción, se plantea si la utilización de las obras para el desarrollo de sistemas de inteligencia artificial requiere la previa autorización de los titulares y el pago de la oportuna licencia. De concluirse que sí, la propiedad intelectual supondría un freno a la innovación y al progreso científico, lo que debe llevar a evaluar la introducción de excepciones y limitaciones al derecho de autor y a los derechos vecinos, conexos o afines para permitir el desarrollo de la inteligencia artificial. En este artículo se estudiarán las soluciones a la cuestión que se han adoptado en la Unión Europea, Estados Unidos y Japón, con el fin de identificar algunas claves que han de ser tenidas en cuenta en la regulación de la materia desde una perspectiva de *lege ferenda*.

PALABRAS CLAVE: Inteligencia artificial, *machine learning*, aprendizaje automático, *deep learning*, aprendizaje profundo, minería de textos y da-

tos, *text data mining*, propiedad intelectual, datos, innovación, excepción, limitación, Unión Europea, UE, Estados Unidos, EEUU, Japón.

SUMARIO:

I. INTRODUCCIÓN. II. QUÉ ES LA INTELIGENCIA ARTIFICIAL Y POR QUÉ EL *MACHINE LEARNING* SUPUSO UN CAMBIO DE PARADIGMA. 1. INTRODUCCIÓN A LA IA. 2. QUÉ ES EL ML Y POR QUÉ IMPLICA UN CAMBIO DE PARADIGMA. 3. ASPECTOS TÉCNICOS DEL ML. 4. ALGUNOS EJEMPLOS PRÁCTICOS PARA ENTENDER LA IMPORTANCIA DE LOS DATOS UTILIZADOS EN EL PROCESO DE ENTRENAMIENTO DE UN MODELO DE ML. III. POR QUÉ EL ENTRENAMIENTO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL ES RELEVANTE PARA LA PROPIEDAD INTELECTUAL. 1. LA ERA DEL *BIG DATA* Y LOS TIPOS DE DATOS GENERADOS. 2. LA CATALOGACIÓN DE LOS «DATOS» COMO SUSCEPTIBLES O NO DE PROTECCIÓN POR LA PROPIEDAD INTELECTUAL. 3. EL ENTRENAMIENTO DE SISTEMAS DE ML COMO ACTIVIDAD RELEVANTE PARA LA PROPIEDAD INTELECTUAL. IV. ESTUDIO DE LAS DISTINTAS APROXIMACIONES A LA CUESTIÓN: UNIÓN EUROPEA, ESTADOS UNIDOS Y JAPÓN. 1. APUNTES PRELIMINARES. 2. UNIÓN EUROPEA. 3. ESTADOS UNIDOS. 4. JAPÓN. V. REFLEXIONES A LA LUZ DE LAS DISTINTAS SOLUCIONES ADOPTADAS EN LA UE, EEUU Y JAPÓN. UNA MIRADA HACIA EL FUTURO DE LA PROPIEDAD INTELECTUAL ANTE LOS RETOS DEL ML. 1. COMPLEJIDAD DE LA CUESTIÓN E INSUFICIENCIA DE LA DICOTOMÍA IDEA-EXPRESIÓN PARA DAR RESPUESTA A LOS RETOS QUE PLANTEA EL ENTRENAMIENTO DEL ML PARA LA PROPIEDAD INTELECTUAL. 2. CASOS DONDE LAS OBRAS Y PRESTACIONES PROTEGIDAS POR DPI SE UTILIZAN EN EL APRENDIZAJE DE LA IA COMO MEDIO PARA ACCEDER A LAS IDEAS O LOS HECHOS. 3. CASOS DONDE LAS OBRAS Y PRESTACIONES PROTEGIDAS POR DPI SE UTILIZAN EN EL APRENDIZAJE DE LA IA CON EL PROPÓSITO DE EXTRAER VALOR DE LOS ELEMENTOS EXPRESIVOS. 4. OTRAS CONSIDERACIONES. VI. CONCLUSIONES. VII. BIBLIOGRAFÍA

TITLE:

Are artificial intelligence systems allowed to «learn» using copyright protected works without the authorization of the owners? The challenges of machine learning examined from a comparative perspective: the EU, the USA and Japan

ABSTRACT:

«Learning» by artificial intelligence systems (in particular, by machine learning systems) poses a significant challenge for intellectual property. The reason for this is that, in most cases, these systems are trained with copyrighted works. Given that in this training process several activities take place that affect the exclusive rights of the owners, mainly the right of reproduction, the question arises of whether using these works to develop ar-

tificial intelligence systems requires the prior authorization of the owners and the payment of a license fee. If the conclusion is that this prior consent is indeed needed, copyright would act as a brake on innovation and scientific progress. In such a context, exceptions and limitations to copyright that allow for the development of artificial intelligence must be assessed. This article will examine the solutions to this question that have been adopted in the European Union, the United States and Japan, in order to identify some key aspects that must be taken into account when regulating this issue from a *lege ferenda* perspective.

KEYWORDS:

Artificial intelligence, machine learning, deep learning, text data mining, intellectual property, copyright, data, innovation, exception, limitation, European Union, United States of America, Japan.

CONTENTS:

I. INTRODUCTION. II. WHAT IS ARTIFICIAL INTELLIGENCE AND WHY MACHINE LEARNING WAS A PARADIGM SHIFT. 1. INTRODUCTION TO AI. 2. WHAT IS ML AND WHY DOES IT IMPLY A PARADIGM SHIFT. 3. TECHNICAL ASPECTS OF ML. 4. SOME PRACTICAL EXAMPLES TO UNDERSTAND THE IMPORTANCE OF THE DATA USED IN THE PROCESS OF TRAINING A ML MODEL. III. WHY TRAINING ARTIFICIAL INTELLIGENCE SYSTEMS IS RELEVANT TO INTELLECTUAL PROPERTY. 1. THE BIG DATA ERA AND THE TYPES OF DATA GENERATED. 2. CLASSIFYING DATA AS COPYRIGHTABLE AND NON-COPYRIGHTABLE. 3. THE RELEVANCE OF ML-SYSTEMS TRAINING TO INTELLECTUAL PROPERTY. IV. ANALYSIS OF THE DIFFERENT APPROACHES TO THE QUESTION POSED: THE EUROPEAN UNION, THE UNITED STATES OF AMERICA AND JAPAN. 1. PRELIMINARY REMARKS. 2. THE EUROPEAN UNION. 3. THE UNITED STATES OF AMERICA. 4. JAPAN. V. REFLECTIONS IN THE LIGHT OF THE DIFFERENT SOLUTIONS ADOPTED IN THE EU, THE USA AND JAPAN. A LOOK INTO THE FUTURE OF INTELLECTUAL PROPERTY IN THE FACE OF THE CHALLENGES POSED BY MACHINE LEARNING. 1. COMPLEXITY OF THE ISSUE AND INSUFFICIENCY OF THE IDEA-EXPRESSION DICHOTOMY TO SOLVE THE CHALLENGES POSED FOR COPYRIGHT BY ML TRAINING. 2. CASES WHERE COPYRIGHTED WORKS AND OTHER SUBJECT MATTER ARE USED IN AI TRAINING AS A WAY TO ACCESS THE IDEAS OR THE FACTS. 3. CASES WHERE COPYRIGHTED WORKS AND OTHER SUBJECT MATTER ARE USED IN AI TRAINING FOR THE PURPOSE OF EXTRACTING VALUE OUT OF THE EXPRESSIVE ELEMENTS. 4. OTHER CONSIDERATIONS. VI. CONCLUSIONS. VII. BIBLIOGRAPHY

I. INTRODUCCIÓN

Amsterdam, 5 de abril de 2016. La expectación era máxima por conocer la que se anunciaba como la última obra del gran maestro Rembrandt, casi cuatro siglos después de su muerte. No se trataba del hallazgo de una obra desaparecida del pintor holandés, sino de una obra nueva, pintada con el sello inconfundible del autor: misma pincelada, composición dinámica, brillante empleo de la técnica del claroscuro. ¿Cómo había sido posible «revivir» a Rembrandt para que pintara una obra más?

El milagro fue posible mediante el empleo de técnicas de *machine learning* («ML»), una rama de la «inteligencia artificial» («IA») que permite que las máquinas aprendan a partir de los datos, sin recibir instrucciones directas¹. El proyecto, denominado «El Nuevo Rembrandt»², nació con un objetivo claro: para revivir a Rembrandt había que crear una máquina que pintara como Rembrandt³. Había que «enseñarle» a pintar como Rembrandt.

Para conseguirlo, 346 obras de Rembrandt fueron escaneadas y analizadas píxel a píxel para crear una extensa base de datos con la que se entrenó a varios algoritmos para identificar y clasificar los patrones utilizados por el maestro holandés para pintar rostros humanos y para dar un sello personal a sus obras en la composición y el uso de luces y las sombras. Otros algoritmos estudiaron las pautas existentes en las capas de pintura de los cuadros. El modelo de ML que se obtuvo era capaz de generar nuevos rasgos faciales replicando el estilo de Rembrandt en todos los aspectos, y «creó» un retrato que, tras ser impreso en 3D, presentaba la misma textura que la de las pinturas del artista. Se logró así el objetivo de «revivir» a Rembrandt mediante la obtención de un modelo de ML que había «aprendido» a pintar como él⁴.

El proyecto «El Nuevo Rembrandt» ilustra bien el avance exponencial que se está produciendo en ML y que afecta a todos los ámbitos: conducir vehículos, diagnosticar enfermedades, detectar el uso fraudulento de tarjetas de crédito, componer canciones o jugar a juegos de mesa son sólo algunos ejemplos de

¹ ¿Qué es el aprendizaje automático? <https://azure.microsoft.com/es-es/overview/what-is-machine-learning-platform/>

² «El Nuevo Rembrandt» fue un proyecto de colaboración entre Microsoft, la entidad financiera ING Bank, la Universidad de Tecnología de Delft y dos museos de arte holandeses (Mauritshuis y Rembrandthuis). Para más información sobre el proyecto, se recomienda la consulta de su página web oficial: <https://www.nextrembrandt.com/>

³ En estos términos lo explicó el director de tecnología del proyecto, Emmanuel Flores: «*Our goal was to make a machine that works like Rembrandt*». La entrevista completa se encuentra disponible en C. BARANIUK, BBC NEWS, 6 de abril de 2016, <https://www.bbc.com/news/technology-35977315>

⁴ Excede del ámbito del presente trabajo el estudio de la originalidad de las obras creadas mediante IA y su protección por derechos de propiedad intelectual, por lo que no se entrará en el análisis de esta cuestión.

tareas que los robots pueden «aprender» a hacer. Las posibilidades son casi infinitas.

La capacidad de aprendizaje de las máquinas mediante ML se encuentra, sin embargo, fuertemente condicionada por un factor: los datos disponibles. Los sistemas de ML, al igual que las personas, aprenden a partir de la experiencia. Sin embargo, en este proceso de aprendizaje, las personas no necesitan disponer de volúmenes de datos tan elevados. Un aprendiz humano no necesita estudiar píxel a píxel 346 obras de Rembrandt para encontrar semejanzas entre ellas. Tampoco necesita ver miles de fotografías de un semáforo para reconocer uno cuando lo ve en la carretera, ni leer miles de textos para aprender a traducir una palabra del inglés al español. Las máquinas, en cambio, son incapaces de aprender si no es a partir de una gran cantidad de datos.

En la era del *big data*, la existencia de un volumen elevado de datos no suele ser un problema, pero sí puede serlo su accesibilidad (principalmente, aunque no solo, en términos de coste), para las compañías que se encargan del desarrollo de sistemas de ML.

En algunas ocasiones, los datos utilizados para el entrenamiento de los sistemas de ML no estarán protegidos por la legislación de propiedad intelectual⁵ —por ejemplo, las obras de Rembrandt pertenecen al dominio público—. En buena parte de los casos, sin embargo, obtener modelos de ML óptimos requerirá entrenar a los sistemas con datos incorporados a materiales protegidos por derechos de propiedad intelectual⁶ («DPI»), lo que implicará, a falta de un límite o excepción que permita la libre utilización de obras protegidas con esa finalidad, obtener una licencia de los titulares de los derechos exclusivos, so pena de abandonar el proyecto por su inviabilidad económica.

Es en este contexto en el que nace la pregunta que da título a este trabajo y que, conforme a autorizada doctrina⁷, puede convertirse en el principal reto de la propiedad intelectual en este siglo: ¿debería la propiedad intelectual permitir que los sistemas de IA sean entrenados con materiales protegidos por DPI sin la autorización de los titulares de los derechos exclusivos? La Organización

⁵ En este artículo utilizaremos la expresión *propiedad intelectual* en el sentido que tiene en el vigente Texto Refundido de la Ley de Propiedad Intelectual española (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia). Esto es, como concepto que comprende el derecho de autor y los derechos afines, vecinos o conexos.

⁶ En consonancia con lo expuesto en la nota anterior, al referirnos en este artículo a derechos de propiedad intelectual (DPI) haremos referencia a los derechos de autor y los derechos afines, vecinos o conexos.

⁷ Así lo entienden M. A. Lemley y B. Casey, que razonan que «*this unique confluence of commercial, normative, and doctrinal factors has teed up what may well be one of the most important legal questions of the coming century: «Will copyright law allow robots to learn?»*». En M. A. LEMLEY, B. CASEY, «Fair Learning», *Texas Law Review*, volumen 99, nº 4, 2021, p. 748.

Mundial de Propiedad Intelectual («OMPI») ha lanzado una consulta pública sobre IA recientemente en la que se trata esta cuestión⁸, respecto de la cual a día de hoy existen todavía más preguntas que respuestas.

Hasta fechas recientes, la doctrina ha prestado poca atención a este tema en comparación con la que ha dedicado al estudio del que es, probablemente, el otro gran reto de la propiedad intelectual en materia de IA en este momento: determinar si las creaciones generadas por IA (como «El Nuevo Rembrandt») pueden considerarse susceptibles de protección jurídica por los DPI —cuestión que excede del objeto de este estudio—. Sin negar la importancia de la materia, no debe tampoco olvidarse que los sistemas de IA no son sólo «creadores» de obras, sino que también son «consumidores»⁹. Y, lo que es más importante: que la calidad de lo que los sistemas de IA son capaces de «crear» está inexorablemente condicionada por la calidad de los datos que «consumen». La máxima *GIGO* (*garbage in, garbage out* o «basura entra, basura sale», en español), frecuentemente utilizada en la jerga del ML, ilustra bien esta realidad: datos de baja calidad utilizados como *input* dan como resultado un *output* de baja calidad. Centrarse sólo en lo que los sistemas de IA «producen» sin atender a los elementos de los cuales depende esa producción (lo que «consumen») supone dejar de lado la mitad del problema.

A la luz de lo anterior, se comprende bien por qué la respuesta a la pregunta que se realiza en el título de este trabajo puede condicionar el desarrollo del ML (y de la IA, como categoría que incluye el ML y otras ramas) en los próximos años. Existen razones económicas de peso para que se permita la utilización de materiales protegidos por DPI en el entrenamiento de sistemas de IA. En lo que se ha acuñado como la Cuarta Revolución Industrial¹⁰, caracterizada precisamente por la importancia estratégica de la IA para la transformación

⁸ WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI), Second Session, *Draft Issues Paper on Intellectual Property and Artificial Intelligence*, 13 de diciembre de 2019, p. 5.

⁹ Hacemos nuestras en este punto las palabras de Duque Lizarralde al recordar, en relación con las labores creativas desarrolladas por IA, que «*los sistemas de IA no son solo productores de arte sino también consumidores*». M. DUQUE LIZARRALDE, «Las obras generadas por inteligencia artificial, un nuevo reto para la propiedad intelectual», *Pe. i. revista de propiedad intelectual*, nº 64, enero-abril 2020, p. 26.

¹⁰ Sobre la Cuarta Revolución Industrial, es cita obligada la obra del mismo nombre del economista alemán Klaus Schwab, fundador y director general del Foro Económico Mundial. Como explica el autor, «*(...) hoy estamos en los albores de una cuarta revolución industrial. Esta comenzó a principios de este siglo y se basa en la revolución digital. Se caracteriza por un internet más ubicuo y móvil, por sensores más pequeños y potentes que son cada vez más baratos, y por la inteligencia artificial y el aprendizaje de la máquina*». En todo caso, como aclara Schwab, la cuarta revolución industrial va más allá de la IA: «*La cuarta revolución industrial, no obstante, no solo consiste en máquinas y sistemas inteligentes y conectados. Su alcance es más amplio. Al mismo tiempo, se producen oleadas de más avances en ámbitos que van desde la secuenciación genética hasta la nanotecnología, y de las energías renovables a la computación cuántica. Es la fusión de estas tecnologías y su interacción a través de los dominios físicos, digitales y biológicos lo que hace que la cuarta revolución industrial sea fundamentalmente diferente de las anteriores*». K. SCHWAB, *La cuarta revolución industrial*, Debate, Barcelona, 2016, pp. 20-21.

económica, parece cada vez más claro que el liderazgo económico (incluso político y militar) de los países estará asociado al liderazgo en IA.

Por ello, al afrontar la cuestión aquí estudiada, los legisladores de los distintos países tendrán que decidir cómo resolver la tensión entre la protección de los DPI y la innovación. Y tendrán que hacerlo, además, con un ojo puesto en lo que ocurre en el resto del mundo, pues a nadie escapa que, en la carrera hacia el primer puesto en IA, las jurisdicciones que ofrezcan condiciones favorables para la innovación (entre ellas, poder utilizar materiales protegidos por DPI sin pagar una licencia a los titulares de los derechos) atraerán talento que impulsará el desarrollo tecnológico. Se da, además, la circunstancia de que en esas jurisdicciones más permisivas, las compañías que desarrollen IA podrán utilizar para el entrenamiento no sólo los materiales protegidos con origen en esa jurisdicción sino también los que tengan origen en otras jurisdicciones —incluyendo los de aquellas que son restrictivas al regular el uso de materiales protegidos para el desarrollo de IA—¹¹. Los países que se decanten por un enfoque protector de los DPI se enfrentan, pues, al riesgo de que los materiales protegidos con origen nacional sean empleados para crear sistemas de ML en las jurisdicciones permisivas —dando lugar a productos y servicios que pueden invadir sus mercados y dejar atrás a sus empresas nacionales— sin que puedan, en contrapartida, permitir a sus empresas nacionales utilizar los materiales protegidos por DPI con origen en esas jurisdicciones permisivas para desarrollar IA. A la vista de lo expuesto, se comprende bien por qué una perspectiva comparada ofrece en esta materia una aproximación particularmente útil al objeto de estudio.

Este artículo se estructurará en diversas secciones. Se comenzará analizando, en la sección II siguiente, qué es la IA desde un punto de vista técnico, y por qué el ML, como rama dentro de la IA, supuso un cambio de paradigma en el modo en que se utilizan los datos para entrenar a los algoritmos. En la sección III se expondrá en qué medida el aprendizaje de los sistemas de IA es relevante para la propiedad intelectual, teniendo en cuenta el tratamiento de los datos incorporados a materiales protegidos por DPI que se efectúa en cada fase de entrenamiento del algoritmo. A continuación, se explicarán en el apartado IV siguiente cuáles son las soluciones a la cuestión que se han planteado en la Unión Europea, Estados Unidos y Japón. Partiendo de ese estudio comparado, en la sección V se expondrán algunas claves que, en atención al estado actual de la cuestión, es esperable que sean tenidas en cuenta desde una perspectiva de

¹¹ Ha de recordarse que, según el principio del trato nacional establecido en el Convenio de Berna para la Protección de las Obras Literarias y Artísticas (el «Convenio de Berna»), los Estados Contratantes deben conceder a las obras cuyo país de origen es otra parte contratante la misma protección que conceden a las obras de sus propios nacionales. De conformidad con el Acuerdo sobre los Aspectos de los Derechos de Propiedad Intelectual relacionados con el Comercio («Acuerdo sobre los ADPIC»), el principio del trato nacional obliga también a los miembros de la Organización Mundial del Comercio que no son parte en el Convenio de Berna.

lege ferenda. En la sección VI, por último, se incluirán unas conclusiones sobre la cuestión estudiada.

II. QUÉ ES LA INTELIGENCIA ARTIFICIAL Y POR QUÉ EL *MACHINE LEARNING* SUPUSO UN CAMBIO DE PARADIGMA

1. INTRODUCCIÓN A LA IA

Para explicar el concepto de IA, habitualmente se recurre a definiciones en las que se atribuye un comportamiento inteligente a los sistemas informáticos, en el sentido de ser sistemas capaces de estudiar su entorno y actuar con cierto grado de autonomía en la consecución de objetivos concretos¹², o se resalta la capacidad de estos sistemas de replicar comportamientos humanos¹³. La IA se utiliza en muchas ocasiones como un término paraguas que engloba distintos significados, pero no existe, como tal, una definición oficial y consensuada sobre lo que es la IA¹⁴.

La necesidad de contar con una definición consensuada sobre la IA para poder iniciar debates sobre el desarrollo de políticas y planes de acción en el seno de la Unión Europea llevó al grupo de expertos de alto nivel sobre IA designado por la Comisión Europea («HLEG», por sus siglas en inglés) a trabajar en una

¹² En estos términos se definía la IA en la Comunicación de la Comisión Europea sobre IA para Europa: «*El término «inteligencia artificial» (IA) se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción —con cierto grado de autonomía— con el fin de alcanzar objetivos específicos. Los sistemas basados en la IA pueden consistir simplemente en un programa informático (p. ej. asistentes de voz, programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), pero la IA también puede estar incorporada en dispositivos de hardware (p. ej. robots avanzados, automóviles autónomos, drones o aplicaciones del internet de las cosas)*». Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones sobre Inteligencia artificial para Europa, Bruselas, 25 de abril de 2018, COM(2018) 237 final.

¹³ En estos términos se define la IA en el trabajo colectivo de Drexel, Hilty y otros autores: «*It is often described as computer-based systems that are developed to mimic human behaviour*». J. DREXL, R. M. HILTY, F. BENEKE, L. DESAUNETTES, M. FICNK, J. GLOBOCNIK, B. GONZÁLEZ OTERO, J. HOFFMAN, L. HOLLANDER, D. KIM, H. RICHTER, S. SCHEUERER, P. R. SLOWINSKI, J. THONEMANN, «Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective», *Max Planck Institute for Innovation & Competition Research Paper No. 19-13*, octubre 2019, p. 3.

¹⁴ Así lo constata el informe del Centro Común de Investigación de la Comisión Europea (más conocido por sus siglas en inglés, JRC) «*AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*». S. SAMOILI, M. LÓPEZ COBO, E. GÓMEZ, G. DE PRATO, F. MARTÍNEZ-PLUMED y B. DELIPETREV, «*AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*», Oficina de Publicaciones de la Unión Europea, Luxemburgo, 2020, p. 4.

definición de IA. En su informe sobre la cuestión¹⁵, el HLEG puso de manifiesto que el concepto de «inteligencia», tanto en máquinas como en humanos, resultaba algo impreciso. Por ello, propuso utilizar en el ámbito de la IA la noción de «racionalidad», entendida como la habilidad para elegir el mejor curso de acción para conseguir un objetivo específico, dados ciertos criterios a optimizar y los recursos disponibles. La definición propuesta por el HLEG¹⁶, que tenía como finalidad clarificar ciertos aspectos de la IA como tecnología y como disciplina científica, es la siguiente:

«Sistemas de software (y posiblemente también de hardware) diseñados por humanos que, ante un objetivo complejo, actúan en la dimensión física o digital percibiendo su entorno a través de la adquisición de datos, interpretando esos datos estructurados o no estructurados, razonando sobre el conocimiento o procesando la información derivada de estos datos y decidiendo las mejores acciones para alcanzar el objetivo marcado. Los sistemas de IA pueden usar reglas simbólicas o aprender un modelo numérico, y también pueden adaptar su comportamiento al analizar cómo el entorno se ve afectado por sus acciones anteriores.

Como disciplina científica, la IA incluye diversos enfoques y técnicas, como el machine learning [aprendizaje automático] (del cual el deep learning [aprendizaje profundo] y el reinforcement learning [aprendizaje reforzado] son ejemplos específicos), el machine reasoning [razonamiento automático] (que incluye planificación, programación, representación y razonamiento del conocimiento, búsqueda y optimización) y la robótica (que incluye control, percepción, sensores y actuadores, así como la integración de todas las demás técnicas en los sistemas ciber-físicos)»¹⁷.

Como ilustra bien la segunda parte de la definición elaborada por el HLEG, la IA como disciplina científica está integrada por distintas ramas. Es importante destacar que no existe consenso entre la comunidad científica acerca de cómo clasificar o dividir esas ramas, debido tanto a los solapamientos existentes entre los distintos enfoques y técnicas como a la rapidez con la que se suceden los avances en el ámbito de la IA.

¹⁵ HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, «A definition of AI: Main capabilities and disciplines», Bruselas, abril 2019.

¹⁶ El primer párrafo de la definición de IA dada por el HLEG ha sido acogida también por el informe del Centro Común de Investigación de la Comisión Europea (JRC) «AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence». El segundo párrafo de la definición del HLEG, en el que se describen sucintamente las ramas que integran la IA es, sin embargo, ampliado y matizado en el informe del Centro Común de Investigación de la Comisión Europea (JRC). *Vid. S. SAMOILI et ál., «AI Watch. Defining Artificial Intelligence...»*, cit., p. 11-17.

¹⁷ Traducción propia. HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, «A definition of AI...», cit., p. 6.

A los efectos de este trabajo, partiremos de la distinción que realiza el HLEG en su definición entre dos grandes grupos de técnicas o subdisciplinas dentro de la IA, en atención a la capacidad de los sistemas de IA para (i) razonar o (ii) para aprender (al margen de la robótica, que el HLEG conceptúa como otra subdisciplina separada).

El primer grupo de técnicas se centra en la capacidad de los sistemas para razonar y tomar decisiones. Para ello, se transforman datos en conocimiento —lo que implica convertir información del mundo real en un lenguaje que sea comprensible para los sistemas— para posteriormente realizar inferencias a partir de reglas simbólicas, de modo que los sistemas puedan tomar decisiones siguiendo un patrón establecido de planificación, búsqueda de soluciones y optimización. En esta categoría se incluiría el *machine reasoning* o razonamiento automático, entre otras subdisciplinas¹⁸.

El segundo grupo de técnicas se refiere a la capacidad de los sistemas para aprender a resolver problemas que no se pueden especificar con precisión o en los que el método de solución no puede ser descrito mediante reglas simbólicas. Por ello, los sistemas son entrenados para resolver problemas mediante la extracción de información a partir de datos¹⁹. En este grupo se enmarcan todas las subdisciplinas que están relacionadas con el aprendizaje de los sistemas de IA, como el ML o el aprendizaje automático, el procesamiento del lenguaje natural (también conocido por sus siglas en inglés, «NLP», *natural language processing*) o la visión artificial o visión por ordenador²⁰.

El presente trabajo se centrará en este segundo grupo de técnicas, en la medida en que, al depender de los datos para el aprendizaje de los sistemas, es donde se plantea la problemática de la utilización de datos protegidos por la propiedad intelectual.

Para terminar esta introducción a la IA, conviene apuntar que no estamos ante una tecnología nueva. La IA comenzó como disciplina académica en la década de 1950. Sin embargo, en su evolución ha alternado períodos de gran desarrollo y creciente interés con otros en los que las expectativas en torno a las posibilidades de la IA han sido menores y, por ende, se le ha dedicado menos atención en el mundo académico y empresarial —hasta el punto de que, contrariamente a lo que ocurre en la actualidad, los investigadores evitaban deliberadamente utilizar el término IA para conseguir financiación para sus proyectos—²¹.

¹⁸ HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, «A definition of AI...», cit., p. 3.

¹⁹ HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, «A definition of AI...», cit., p. 3.

²⁰ S. SAMOILI *et ál.*, «AI Watch. Defining Artificial Intelligence...», cit., p. 11.

²¹ B. DELIPETREV, C. TSINARAKLI, U. KOSTIC, «Historical Evolution of Artificial Intelligence», Oficina de Publicaciones de la Unión Europea, Luxemburgo, 2020, p. 10.

Desde 1990, la IA ha experimentado un desarrollo exponencial que continúa en la actualidad y que la ha situado como «*una de las tecnologías más estratégicas del siglo XXI*»²², llamada a desempeñar un papel central en la Cuarta Revolución Industrial²³. Entre los factores que han impulsado el desarrollo de la IA cabe destacar tres²⁴:

- i. El aumento en la cantidad de datos disponible.
- ii. El incremento en la potencia de computación y en la capacidad de almacenamiento.
- iii. El desarrollo de nuevos algoritmos y métodos de ML —cuestión que se tratará en el siguiente apartado—.

2. QUÉ ES EL ML Y POR QUÉ IMPLICA UN CAMBIO DE PARADIGMA

El ML o aprendizaje automático es, como se ha anticipado, una de las ramas de la IA. Puede definirse como el estudio científico de los algoritmos que los sistemas informáticos aprenden mediante la experiencia²⁵. Consiste en entrenar a un programa informático a identificar patrones en los datos disponibles y aplicar el conocimiento a nuevos datos²⁶.

El ML supuso un cambio de paradigma respecto de la programación clásica, marcando un hito en la historia de la IA y propiciando su evolución desde 1990 y, en particular, en la última década con el desarrollo del *deep learning* o aprendizaje profundo²⁷, una de las ramas del ML que más ha progresado. Para evitar confusiones terminológicas, en este trabajo nos referiremos de forma genérica a las técnicas de ML sin hacer distinciones entre sus distintas ramas, pero han de entenderse incluidas todas ellas, y en particular el *deep learning* o aprendizaje profundo como subdisciplina del ML. Del mismo modo, las referencias a la IA han de entenderse también efectuadas a todas sus ramas, entre las que se incluye, como se ha expuesto, el ML.

Para explicar por qué el ML es concebido como un cambio de paradigma en el ámbito de la IA, es preciso explicar brevemente el funcionamiento de la

²² Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones sobre Inteligencia artificial para Europa, Bruselas, 25 de abril de 2018, COM(2018) 237 final, p.2.

²³ K. SCHWAB, *La cuarta revolución industrial*, Debate, Barcelona, 2016, pp. 20-21.

²⁴ GOBIERNO DE ESPAÑA. «Estrategia Nacional de Inteligencia Artificial. Versión 1.0», noviembre 2020, p.6. <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf>

²⁵ S. SAMOILI *et ál.*, «AI Watch. Defining Artificial Intelligence...», cit., p. 6.

²⁶ J. DREXL *et ál.*, «Technical Aspects of Artificial Intelligence...», cit., p. 3.

²⁷ El *deep learning* o aprendizaje profundo es una rama del ML (que es a su vez, como se ha visto, una rama de la IA). El *deep learning* se diferencia del ML en cómo aprende el algoritmo y en cuántos datos utiliza cada tipo de algoritmo. Para más información sobre las diferencias entre estos conceptos, *vid.* IBM CLOUD EDUCATION, «AI vs. Machine Learning vd. Deep Learning vs. Neural Networks: What's the Difference?», 27 mayo 2020. <https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>

programación clásica. Esta última consistía en definir un conjunto explícito de instrucciones para realizar una tarea en particular, en lenguajes de programación como Java o C. Estas reglas que especificaba el programador eran combinadas con datos para generar respuestas²⁸.

Un ejemplo de un programa de IA de adopción temprana es el ajedrez por ordenador. Para su desarrollo, los programadores proporcionaban al ordenador un conjunto de reglas preconcebidas sobre cómo jugar al ajedrez y, a partir de esas reglas, el sistema era capaz de jugar por sí mismo. Uno de los grandes hitos en la evolución de la IA se produjo cuando en el año 1997 un ordenador llamado Deep Blue, desarrollado por IBM, fue capaz de vencer a la leyenda del ajedrez Garry Kasparov. La hazaña, que vista más de veinte años después puede no impresionar demasiado, asombró al mundo en aquel entonces, máxime teniendo en cuenta que Kasparov había vencido a Deep Blue sólo un año antes²⁹.

El enfoque seguido en la programación clásica, sin embargo, no daba buenos resultados en todos los ámbitos. Los investigadores en IA se dieron cuenta de que en algunas tareas la formulación de reglas o instrucciones resultaba muy complicada, lo que daba como resultado que los sistemas obtuvieran un rendimiento deficiente. Un ejemplo se puede encontrar en el ámbito de la percepción sensorial, donde resulta difícil definir instrucciones que permitan a los sistemas identificar imágenes.

El ML proporcionó una nueva aproximación para resolver estos problemas. En lugar de programar un sistema con un conjunto de reglas preconcebidas para realizar una tarea específica, al sistema se le proporcionan datos y respuestas a los problemas planteados para que sea capaz de extraer las reglas³⁰. En otras palabras: en el ML los algoritmos no reciben instrucciones específicas para el aprendizaje, sino que «aprenden» ellos mismos de la experiencia a partir de un conjunto de datos, siendo posteriormente capaces de realizar predicciones sobre nuevos conjuntos de datos.

El cambio de paradigma que supuso el ML puede ilustrarse mejor acudiendo al ejemplo —comúnmente utilizado— de cómo distinguir un gato en una imagen. En la programación clásica los algoritmos habrían sido instruidos con una serie de reglas sobre cómo distinguir un gato y procesarían la imagen tratando de

²⁸ B. DELIPETREV, C. TSINARAKLI, U. KOSTIC, «Historical Evolution of Artificial Intelligence», cit., p. 11.

²⁹ HISTORY.COM EDITORS, «Deep Blue defeats Garry Kasparov in chess match», 16 noviembre 2009. <https://www.history.com/this-day-in-history/deep-blue-defeats-garry-kasparov-in-chess-match#:~:text=1997%20May%2011-,Deep%20Blue%20defeats%20Garry%20Kasparov%20in%20chess%20match,to%20one%2C%20with%20three%20draws.>

³⁰ B. DELIPETREV, C. TSINARAKLI, U. KOSTIC, «Historical Evolution of Artificial Intelligence», cit., p. 11.

identificar los ojos, las orejas, etc. En ML, el algoritmo aprendería a reconocer un gato tras ser entrenado con múltiples imágenes de gatos³¹ —de forma similar a como aprenden los humanos, esto es, mediante la experiencia—.

El ámbito de la percepción no es el único en el que las técnicas de ML resultan útiles, sino también, en general, en todas aquellas tareas que resultan difíciles de definir y no pueden describirse de manera exhaustiva mediante reglas de comportamiento simbólicas³².

Otro ejemplo de tarea donde la definición de reglas preconcebidas no daba buenos resultados y, por ende, la programación clásica fracasaba, era otro juego de mesa: el Go, un juego ancestral originario de China considerado de mayor dificultad que el ajedrez por el mayor número de movimientos posibles en cada jugada. En Go, el número de configuraciones posibles de juego es de 10 elevado a la potencia de 170 —mayor que el número de átomos en el universo—³³. Esta complejidad hacía muy difícil la formulación de reglas, lo que daba como resultado que los programas de Go obtuvieran malos resultados frente a jugadores humanos. A pesar de que, como se ha indicado, ya en 1997 el programa Deep Blue había sido capaz de vencer al campeón de ajedrez Garry Kasparov, más de una década después aún parecía imposible, según los expertos, que un programa de ordenador fuera capaz de emular la gesta en el caso de Go.

Es en este contexto en el que, en el año 2015, tuvo lugar uno de los grandes hitos del ML: por primera vez en la historia, un programa de ordenador llamado AlphaGo logró vencer a un jugador profesional de Go³⁴. El éxito fue todavía mayor cuando Alpha Go ganó en 2016 a Lee Sedol, el jugador número 1 del mundo³⁵.

El éxito de Alpha Go se encuentra en el uso de las técnicas de ML y en el proceso de entrenamiento seguido por los algoritmos. Alpha Go aprendió a partir de millones de movimientos de jugadores humanos y de jugar miles de partidas contra sí misma, como explican sus creadores:

«Primero entrenamos la red en 30 millones de jugadas de partidas jugadas por humanos expertos, hasta que pudo predecir el movimiento humano el 57%

³¹ B. DELIPETREV, C. TSINARAKLI, U. KOSTIC, «Historical Evolution of Artificial Intelligence», cit., p. 11.

³² HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, «A definition of AI...», cit., p. 3.

³³ *Vid.* página web oficial de AlphaGo: <https://deepmind.com/research/case-studies/alphago-the-story-so-far>

³⁴ BBC NEWS, «Google achieves AI 'breakthrough' by beating Go champion», 27 enero 2016. <https://www.bbc.com/news/technology-35420579>

³⁵ BBC NEWS, «Google AI wins second Go game against top player», 10 marzo 2016. <https://www.bbc.com/news/technology-35771705>

*del tiempo (el récord anterior a AlphaGo era de 44%). Pero nuestro objetivo es vencer a los mejores jugadores humanos, no solo imitarlos. Para hacerlo, AlphaGo aprendió a descubrir nuevas estrategias por sí misma, jugando miles de juegos entre sus redes neuronales y mejorándolas gradualmente mediante un proceso de prueba y error conocido como aprendizaje por refuerzo (...))*³⁶.

Interesa destacar que, en sus partidas frente a jugadores humanos, Alpha Go utilizó estrategias no convencionales y creó nuevas jugadas, lo que llevó al campeón mundial de Go, Lee Sedol, a hacer las siguientes declaraciones: «*Pensaba que AlphaGo estaba basada en cálculo probabilístico y que era solamente una máquina. Pero cuando vi esta jugada, cambié de opinión. Sin duda, AlphaGo es creativa*»³⁷.

Esta capacidad de los sistemas de ML para tomar decisiones de forma autónoma una vez que han sido entrenados, con un componente de «creatividad» en el resultado final, es un factor a considerar en el debate sobre si las creaciones generadas por IA, como «El Nuevo Rembrandt», pueden considerarse susceptibles de protección por DPI. En el caso de «El Nuevo Rembrandt», fue el sistema de ML el que eligió los rasgos faciales del personaje del retrato a partir de los datos de entrenamiento utilizados.

En conclusión, el ML supuso un cambio de paradigma en la IA, en la medida en que se pasó de «enseñar» a los sistemas mediante un conjunto de reglas preconcebidas a dejar que fueran los propios sistemas los que «aprendieran» a partir de un conjunto de datos y sin instrucciones específicas. Esta nueva forma de desarrollar IA supuso, respecto de la programación clásica, un cambio en los *inputs* utilizados para el entrenamiento —se precisan ahora grandes cantidades de datos— y en los *outputs* obtenidos —sistemas que actúan con un cierto grado de autonomía y que toman decisiones que no son siempre fácilmente explicables—.

3. ASPECTOS TÉCNICOS DEL ML

Desde un punto de vista técnico, el ML consiste en el análisis de datos mediante algoritmos³⁸. El proceso consta de diferentes etapas, que se describirán muy brevemente en esta sección³⁹.

³⁶ Traducción propia. En D. SILVER, D. HASSABIS, «AlphaGo: Mastering the ancient game of Go with Machine Learning», *Google AI Blog*, 27 enero 2016. <https://ai.googleblog.com/2016/01/alphago-mastering-ancient-game-of-go.html>

³⁷ *Vid.* página web oficial de AlphaGo: <https://deepmind.com/research/case-studies/alphago-the-story-so-far>

³⁸ J. DREXL *et ál.*, «Technical Aspects of Artificial Intelligence...», op. cit., p. 4.

³⁹ La descripción que aquí se ofrece sobre las fases de entrenamiento del ML supone una simplificación de un proceso que es enormemente complejo, pero cuyo análisis en profundidad no se considera necesario a efectos de este trabajo. Para un estudio más detallado sobre esta cuestión, *vid.* J. DREXL *et ál.*, «Technical Aspects of Artificial Intelligence...», op. cit. También D. LEHR y

Imaginemos que queremos crear un sistema de IA que sea capaz de distinguir imágenes de perros y gatos. En el lenguaje de ML, este sistema se denomina «modelo», y se obtiene como resultado de un proceso de entrenamiento. Para entrenar al modelo, se necesita recopilar grandes cantidades de datos: es el punto de partida para el ML.

A continuación se describirán sucintamente las distintas fases que deben seguirse para obtener un modelo de ML⁴⁰:

1. Recopilación de datos

Tras identificar el problema a resolver, el proceso comienza con la recopilación de datos suficientes⁴¹ para entrenar al modelo. La importancia de este paso es primordial, ya que la calidad y cantidad de los datos que se recopilen determinará el rendimiento del modelo y su capacidad para hacer predicciones precisas⁴². En nuestro ejemplo, los datos recopilados serían imágenes de perros y gatos (preferiblemente, varios miles de cada tipo).

2. Preparación de datos

El siguiente paso es cargar los datos recopilados en una plataforma digital adecuada para el proceso de entrenamiento. En ocasiones, los datos recopilados necesitan pasar por una fase de preparación, que implica realizar ajustes o modificaciones en los datos (por ejemplo, eliminar duplicidades, corregir errores, normalizar los datos, etc.). Además, en esta fase el conjunto de datos debe dividirse en dos grandes bloques: los datos de entrenamiento (que constituirán la mayor parte del conjunto, y que se utilizarán para entrenar al modelo) y los datos de evaluación (que constituirán una parte pequeña del conjunto, que se empleará para evaluar el desempeño predictivo del modelo)⁴³.

P. OHM, «Playing with the Data: Why Legal Scholars Should Learn About Machine Learning», *University of California Davis Law Review*, 51, 2017, pp. 655-717.

⁴⁰ En la exposición seguiremos el modelo de siete pasos descrito en el vídeo de Google Cloud AI Adventures protagonizado por Y. GUO denominado *7 Steps of Machine Learning*, cuyo visionado se recomienda para una mejor comprensión sobre este punto. <https://www.youtube.com/watch?v=nKW8Ndu7Mjw&list=PLIivdWyY5sqJxnwJhe3etaK7utrBiPBQ2&t=25s> Este modelo de siete pasos también se describe en Y. GUO, *The 7 Steps of Machine Learning*, 31 agosto 2017. <https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e>

⁴¹ Como indican D. LEHR y P. OHM, «Playing with the Data...», op. cit., pp. 679, resulta difícil establecer una regla sobre cuál es el número de datos necesario en la práctica. Con frecuencia, disponer de unos cientos o unos miles puede resultar insuficiente para algunas aplicaciones, requiriéndose decenas de miles. En ocasiones, el número óptimo podrá ser de cientos de miles de datos o incluso millones.

⁴² Y. GUO, *The 7 Steps of Machine Learning*, 31 agosto 2017. <https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e>

⁴³ Y. GUO, *The 7 Steps...*, op. cit.

3. Elección de un modelo

Después de recopilar y preparar los datos, es el momento de elegir qué tipo de modelo se empleará, para lo cual los científicos de datos atienden, entre otros parámetros, al tipo de datos utilizado (texto, imágenes, música, etc.)⁴⁴. Se trata de una cuestión compleja cuyo estudio detallado no se precisa a los efectos del presente trabajo.

4. Entrenamiento

Durante esta fase, un algoritmo de ML aprende a partir de los datos de entrenamiento a identificar reglas que le permitan efectuar predicciones. Como resultado del entrenamiento se obtiene un modelo de ML, que puede definirse como un algoritmo basado en una función matemática (no lineal) que genera resultados basados en los patrones aprendidos en los datos de entrenamiento⁴⁵. En nuestro ejemplo, el algoritmo aprendería a partir de las miles de imágenes de gatos y perros incluidas en el conjunto de datos de entrenamiento, con el fin de identificar patrones para distinguirlos.

Atendiendo al tipo de datos y a la forma en que se utilizan en el proceso de aprendizaje del algoritmo, pueden distinguirse tres métodos de entrenamiento: aprendizaje supervisado⁴⁶, aprendizaje no supervisado⁴⁷ y aprendizaje reforzado⁴⁸.

5. Evaluación

Una vez que el modelo ha sido entrenado, se debe someter a pruebas para evaluar su desempeño predictivo. Para ello, el modelo se aplica a nuevos datos que no han sido utilizados en el proceso de entrenamiento: los datos de evaluación. Se busca, con este ejercicio, obtener una representación

⁴⁴ Y. GUO, *The 7 Steps...*, op. cit.

⁴⁵ J. DREXL *et ál.*, «Technical Aspects of Artificial Intelligence...», op. cit., p. 6.

⁴⁶ En el aprendizaje supervisado, los datos de entrenamiento se etiquetan, lo que permite al modelo conocer si la predicción que efectuó en el proceso de optimización fue correcta o incorrecta. Es la forma más utilizada de ML en la actualidad. J. DREXL *et ál.*, «Technical Aspects of Artificial Intelligence...», op. cit., pp. 4, 13.

⁴⁷ En el aprendizaje no supervisado, los datos de entrenamiento no están etiquetados. Por ello, el modelo es entrenado para identificar similitudes, paralelismos o diferencias en los datos. J. DREXL *et ál.*, «Technical Aspects of Artificial Intelligence...», op. cit., p. 13.

⁴⁸ En el aprendizaje reforzado o aprendizaje por refuerzo, el modelo no se entrena tomando como base un conjunto de datos preexistente, sino que recopila datos a través de simulaciones o de juegos, donde el algoritmo descubre las reglas a partir de una retroalimentación continua respecto de las acciones que realiza durante el aprendizaje. J. DREXL *et ál.*, «Technical Aspects of Artificial Intelligence...», op. cit., p. 13.

acerca de cómo funcionaría el modelo en la vida real⁴⁹. En nuestro ejemplo, aplicaríamos el modelo de ML a imágenes de gatos y perros incluidas en el conjunto de datos de evaluación para medir la precisión al distinguir imágenes de cada tipo.

6. *Ajuste de hiperparámetros*

Tras la fase de evaluación, puede revelarse la necesidad de mejorar el desempeño predictivo del modelo ajustando algunos parámetros. Estos ajustes son, de hecho, muy habituales en la práctica, y rara es la vez que se obtiene un modelo que realice predicciones con una precisión suficiente sin haber dedicado algún tiempo a realizar ajustes (se habla por ello, en ocasiones, de que el ML es un proceso de prueba y error, o de que es «más arte que ciencia»)⁵⁰. Un ejemplo de estos ajustes es el referido al número de iteraciones en los datos de entrenamiento (esto es, el número de veces que el modelo es entrenado con el conjunto de datos de entrenamiento). En ocasiones, al «mostrar» los datos al modelo más de una vez, el modelo gana en precisión⁵¹. No nos extenderemos en el análisis del resto de parámetros que pueden ajustarse en esta fase, al ser una cuestión compleja respecto de la que no es preciso profundizar a los efectos del estudio que aquí se realiza.

7. *Implementación del modelo de ML (predicción)*

El propósito del proceso de ML es obtener un modelo que sea capaz de responder preguntas o hacer predicciones sobre nuevos datos. Tras las diversas fases de desarrollo que se han descrito, el modelo estará preparado para hacer predicciones en el mundo real. En el ejemplo dado, nuestro modelo de ML debería ser capaz de predecir, con una tasa de precisión aceptable, si una imagen nueva que se le presenta corresponde a un gato o a un perro.

Es importante señalar que, una vez que se obtiene un modelo de ML siguiendo el proceso de entrenamiento descrito, el modelo es completamente utilizable sin los datos de entrenamiento y sin los datos de evaluación. Esto implica que el desarrollador de un modelo de ML puede venderlo con fines comerciales o no comerciales sin necesidad de revelar los datos utilizados para el entrenamiento del modelo⁵².

⁴⁹ Y. GUO, *The 7 Steps...*, op. cit.

⁵⁰ D. LEHR y P. OHM, «Playing with the Data...», op. cit., pp. 699-700.

⁵¹ Y. GUO, *The 7 Steps...*, op. cit.

⁵² J. DREXL *et ál.*, «Technical Aspects of Artificial Intelligence...», op. cit., p. 8.

4. ALGUNOS EJEMPLOS PRÁCTICOS PARA ENTENDER LA IMPORTANCIA DE LOS DATOS UTILIZADOS EN EL PROCESO DE ENTRENAMIENTO DE UN MODELO DE ML

Como se ha avanzado en el apartado anterior, los datos utilizados en el proceso de entrenamiento del modelo de ML tienen una importancia capital. Cuanto mejor sea el conjunto de datos de entrenamiento en términos de calidad, cantidad y variedad, más preciso será el modelo de ML⁵³ en sus predicciones. Y al contrario: un conjunto de datos de entrenamiento donde la calidad, cantidad y variedad sean bajas dará como resultado un modelo de ML con un desempeño predictivo deficiente. Los científicos de datos se refieren habitualmente a esta correlación utilizando la gráfica expresión *GIGO* (*garbage in, garbage out*).

Lo anterior se explica porque los modelos de ML no aprenden a razonar por sí mismos a partir de los datos con los que son entrenados, sino que realizan sus predicciones a partir de cálculos probabilísticos. Si el conjunto de datos de entrenamiento no está bien confeccionado en atención al problema que se pretende resolver mediante el modelo de ML, las predicciones que realice el modelo no serán de utilidad.

Para ilustrarlo con un ejemplo, pensemos en un modelo de ML que componga canciones y que sea entrenado exclusivamente con canciones pertenecientes al dominio público. Ese modelo de ML difícilmente podrá componer canciones en estilos musicales desarrollados en los últimos veinte años, como el *reggaeton* o el *trap*, en la medida en que no los «conoce». Una situación similar ocurriría con un traductor de textos inglés-español que fuera entrenado con novelas pertenecientes al dominio público: sería incapaz de traducir expresiones del lenguaje moderno, por lo que probablemente no funcionaría correctamente al traducir el contenido de una publicación en una red social⁵⁴.

Otro ejemplo de la correlación entre el conjunto de datos de entrenamiento y el desempeño del modelo de ML es, como apunta Sobel⁵⁵, el caso de la función llamada «Respuesta Inteligente» (*Smart Reply*) que agregó Google a sus servicios de correo electrónico. La función «Respuesta Inteligente» es capaz de

⁵³ J. DREXL *et ál.*, «Technical Aspects of Artificial Intelligence...», op. cit., p. 8.

⁵⁴ Como se puede constatar a partir de estos ejemplos, la utilización de obras del dominio público para el entrenamiento de sistemas de ML, si bien resuelve el problema de la tensión entre la protección de los DPI y la innovación desde la perspectiva legal, no es una buena solución desde el punto de vista tecnológico, pues los modelos de ML entrenados únicamente con esas obras pueden tener un desempeño predictivo deficiente en determinados contextos. Por esta razón, no compartimos plenamente la opinión de Chiou sobre el uso de obras del dominio público para el aprendizaje de ML como posible solución para evitar el pago de licencias a los titulares de los DPI, en tanto que normalmente ello dará lugar a modelos de ML de calidad inferior. T. CHIOU, «Copyright Law and Algorithmic Creativity: Monopolizing Inspiration?», *EU Internet Law in the Digital Single Market* (Ed. T. Synodinou, P. Jougleux, C. Markou, T. Prastitou-Merdi), Springer, Cham, 2021, p. 283.

⁵⁵ B. L. W. SOBEL, «Artificial Intelligence's fair use crisis», *Columbia Journal of Law & The Arts*, 41, 2017, p. 22.

generar de modo automático respuestas para los correos que los usuarios reciben, atendiendo al contexto de la conversación⁵⁶. Para conseguirlo, el algoritmo de «Respuesta Inteligente» fue inicialmente entrenado a partir de un conjunto de datos compuesto por 239 millones de correos electrónicos⁵⁷. Ocurrió, sin embargo, que como el estilo de redacción de esos correos electrónicos no era muy sofisticado, el del algoritmo tampoco lo era⁵⁸. Para mejorar el estilo de redacción del algoritmo, sus desarrolladores decidieron entrenarlo con algunas novelas⁵⁹.

III. POR QUÉ EL ENTRENAMIENTO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL ES RELEVANTE PARA LA PROPIEDAD INTELECTUAL

1. LA ERA DEL *BIG DATA* Y LOS TIPOS DE DATOS GENERADOS

Como se ha visto en la sección anterior, los datos son el elemento más importante del proceso de entrenamiento de un sistema de ML. No es, por ello, una coincidencia que el ML haya alcanzado su punto álgido de desarrollo en la era del *big data*: la gran cantidad de datos existentes ha sido, junto con un aumento exponencial en la potencia informática, la clave para explicar el auge de los modelos de ML.

Hoy hay más datos que nunca y las cifras siguen creciendo a medida que aumenta el número de usuarios de Internet de forma significativa año tras año y se acelera la digitalización —todavía a ritmo mayor por efecto del COVID-19—. Según el informe *Data Never Sleeps 8.0*, que ofrece algunas mediciones de los datos que son generados durante cada minuto en el mundo a través de las plataformas y redes sociales más populares, en abril de 2020 Internet llegó al 59% de la población mundial, esto es, a 4.570 millones de personas (lo que supone un aumento del 6% desde enero de 2019)⁶⁰. Las cifras de los datos generados por minuto que se presentan en ese informe sirven bien para demostrar por qué se considera que estamos en la era de los datos. Por poner algunos ejemplos, según el informe, en el año 2020 durante cada minuto del día, como media, los usuarios de WhatsApp intercambiaron 46.666.667 mensajes, los usuarios de Instagram publicaron 347.222 historias, los usuarios de Facebook subieron a la red 147.000 fotos e intercambiaron 150.000 mensajes, los usuarios de YouTube cargaron 500 horas de vídeo y

⁵⁶ Así se explica en la página web de Smart Reply. <https://developers.google.com/ml-kit/language/smart-reply>

⁵⁷ B. L. W. SOBEL, «Artificial Intelligence's», op. cit., p. 22.

⁵⁸ B. L. W. SOBEL, «Artificial Intelligence's...», op. cit., p. 22.

⁵⁹ B. L. W. SOBEL, «Artificial Intelligence's...», op. cit., p. 22.

⁶⁰ DOMO, *Data Never Sleeps 8.0*. <https://web-assets.domo.com/blog/wp-content/uploads/2020/08/20-data-never-sleeps-8-final-01-Resize.jpg>

319 personas se hicieron usuarios de Twitter⁶¹ (plataforma donde, de media, cada minuto se publican más de 500.000 *tweets*⁶²).

La abundancia de datos generados por personas no sólo encuentra su origen, como es lógico, en plataformas de colaboración y redes sociales. En la llamada sociedad del conocimiento, la generación y el intercambio de conocimiento e información es constante. Por poner otro ejemplo, la comunidad científica global genera 1,5 millones de artículos doctrinales nuevos cada año, o aproximadamente un nuevo artículo cada 30 segundos, como apuntan Margoni y Martin⁶³.

Pero los datos que se generan no tienen su origen únicamente en la creación activa de contenidos por parte de personas. Existen, en la actualidad, muchas otras fuentes de producción de datos, como los que se comparten «entre máquinas» (los termómetros de las ciudades, los GPS de vehículos, etc.), los que proceden de fuentes biométricas (sensores de huellas dactilares, de reconocimiento de voz, etc.), los que generan los usuarios de Internet al navegar por la red, los que se generan a partir de transacciones económicas (efectuar reservas de billetes de avión, añadir artículos a un carrito de compra virtual de una página web, etc.)⁶⁴, y un largo etcétera.

Esta gran cantidad de datos que se crea minuto a minuto a nivel global brinda infinitas oportunidades para la investigación en el ámbito del ML. Ha de tenerse en cuenta, además, que aunque algunos de estos datos no son accesibles (por ejemplo, los que se generan en el ámbito de una empresa y no son compartidos), existen otros muchos que son plenamente accesibles a través de Internet. Lo anterior implica que, con cierta frecuencia, un científico de datos se encontrará a tan solo un *clic* de distancia de acceder a datos suficientes como para conformar un conjunto de datos de entrenamiento con el que desarrollar un modelo de ML con una precisión aceptable⁶⁵.

El problema es que, en no pocas ocasiones, esos mismos datos que, desde el punto de vista tecnológico se presentarían como accesibles e idóneos para el aprendizaje de un modelo de ML, podrían estar protegidos por la propiedad

⁶¹ DOMO, *Data Never Sleeps 8.0*. <https://web-assets.domo.com/blog/wp-content/uploads/2020/08/20-data-never-sleeps-8-final-01-Resize.jpg>

⁶² DOMO, *Data Never Sleeps 7.0*. <https://web-assets.domo.com/blog/wp-content/uploads/2019/07/data-never-sleeps-7-896kb.jpg>

⁶³ M. KRETSCHMER, T. MARGONI, «The Text and Data Mining exception in the Proposal for a Directive on Copyright in the Digital Single Market: Why it is not what EU copyright law needs», 25 abril 2018, p. 1.

⁶⁴ ¿Qué es big data? <https://www.masterbigdataucm.com/que-es-big-data/>

⁶⁵ En el sencillo ejemplo utilizado en la sección anterior, donde se obtenía un modelo de ML capaz de distinguir entre imágenes de gatos y de perros, los datos de entrenamiento podrían fácilmente obtenerse con tan solo teclear «gato» y «perro» en el buscador Google y acceder a la categoría de imágenes. Cuestión distinta será si utilizar esos datos puede infringir DPI de sus titulares, que es lo que se estudiará en este artículo.

intelectual⁶⁶. Ello no ocurrirá con los datos generados por los GPS de los vehículos o por los sensores de huellas dactilares (por volver sobre algunos de los ejemplos anteriores), pero sí será generalmente el caso de las publicaciones científicas, y podrá suceder también —en función de la legislación aplicable y del caso concreto— con una fotografía compartida en Instagram, un poema publicado en Facebook, un *tweet* que contenga unas líneas de denuncia social en estilo satírico o un video de YouTube de un artista que se grabe mientras canta. Esto obliga a las empresas de IA a adoptar precauciones en función de los datos utilizados, como se expondrá a continuación.

2. LA CATALOGACIÓN DE LOS «DATOS» COMO SUSCEPTIBLES O NO DE PROTECCIÓN POR LA PROPIEDAD INTELECTUAL

Varios de los ejemplos de «datos» que se han citado en el apartado anterior podrían estar incorporados a obras protegidas por DPI. En la medida en que, como se ha explicado anteriormente, el proceso de entrenamiento de un sistema de ML implica copiar los datos y, en ocasiones, modificarlos —particularmente, durante la fase de preparación de datos—, parece claro que, si en el conjunto de datos de entrenamiento se incluyen obras protegidas por la propiedad intelectual, se podrían infringir los derechos exclusivos de los titulares.

Nótese que la conclusión que se alcance sobre si cada una de las miles de obras utilizadas para entrenar a un sistema de ML, individualmente considerada, puede estar protegida por DPI bajo la legislación aplicable no es relevante a efectos de responder a la pregunta que se plantea en el título de este trabajo. O dicho de otro modo, no arroja luz, a efectos del análisis, evaluar hasta qué punto es probable que todos o solo algunos de los *tweet* que se utilicen para entrenar a un sistema de ML que aprenda a redactar textos puedan ser considerados originales bajo el estándar de la legislación que resulte aplicable. Es el mero riesgo de que todos o algunos de esos *tweets* puedan serlo y, por ende, de que su utilización para entrenar a un sistema de ML sin la previa autorización del titular constituya una infracción de los DPI, lo que determina la necesidad de plantearse el encaje de esa actividad en la legislación de propiedad intelectual aplicable.

La razón es puramente práctica. Si en un proyecto de desarrollo de un modelo de ML que redactara textos para ser publicados en Twitter tuviera que incluirse el coste de analizar, uno a uno en un conjunto de varios miles, si cada *tweet* utilizado cumple o no *a priori* con el estándar de originalidad exigido por la legislación aplicable, el proyecto sería económicamente inviable. Por ello, las empresas que desarrollen modelos de ML podrán, a lo sumo, clasificar los

⁶⁶ Los datos podrán estar también protegidos por la normativa de protección de datos personales. Se trata de una cuestión cuyo análisis excede el alcance de este estudio.

«datos» que utilizan para el entrenamiento en *no susceptibles de protección* por DPI (por pertenecer al dominio público o por no ser una creación literaria, artística o científica ni una prestación protegida por los derechos vecinos, afines o conexos) y en *susceptibles de protección* por DPI. Si llegan a la conclusión de que los «datos» están incorporados a objetos que son susceptibles de protección, deberán estudiar de qué manera la legislación de propiedad intelectual aplicable trata la cuestión antes de utilizar esos datos para el desarrollo de sistemas de ML, so pena de exponerse a reclamaciones de los titulares de los derechos exclusivos, que pueden alcanzar cuantías millonarias por el elevado volumen de datos empleado.

En esa aproximación al problema que realicen las empresas que desarrolleen IA cabe plantearse si es posible sostener que en el entrenamiento del sistema de ML no se está realizando una utilización integral del objeto protegido por DPI, sino tan solo de elementos aislados y desagregados. Aunque sería una línea argumental que se podría explorar en determinados casos, el potencial número de reclamaciones al que pueden enfrentarse los desarrolladores de IA impone la adopción de un enfoque prudente⁶⁷.

3. EL ENTRENAMIENTO DE SISTEMAS DE ML COMO ACTIVIDAD RELEVANTE PARA LA PROPIEDAD INTELECTUAL

En esta sección se estudiará si el entrenamiento de sistemas de ML debe ser considerado como una actividad relevante a efectos de la propiedad intelectual. Para ello, se expondrá en primer lugar de qué manera se pueden ver afectados los derechos exclusivos de los titulares de DPI en cada una de las fases de desarrollo de un modelo de ML⁶⁸.

Las fases que comprende la obtención de un modelo de ML se han analizado en el apartado II.3 de este trabajo. Como allí se ha expuesto, el proceso comienza con la recopilación de datos, que deben ser posteriormente incorporados a una

⁶⁷ La estrategia de defender que se están utilizando datos de forma desagregada puede verse, además, amenazada por precedentes jurisprudenciales como la Sentencia del Tribunal de Justicia de la Unión Europea («TJUE») de 16 de julio de 2009 dictada en el caso *Infopaq International A/S v. Danske Dagblades Forening* (asunto C-5/08). En este asunto, el TJUE concluyó que un extracto de once palabras es considerado una reproducción parcial en el sentido del artículo 2 de la Directiva 2001/29/CE del Parlamento Europeo y el 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 («Directiva 2001/29/CE»), si el citado extracto contiene algún elemento capaz de expresar la creación intelectual propia del autor.

⁶⁸ En la medida en que en la sección IV siguiente se expondrán las distintas aproximaciones a la cuestión estudiada desde la perspectiva comparada (se verán, en particular, los casos de la Unión Europea, Estados Unidos y Japón), no se hará referencia en este apartado a ninguna jurisdicción en particular, sino que se hará mención a las facultades exclusivas de los titulares de DPI y a su posible afectación durante el proceso de ML partiendo de principios comúnmente aceptados.

plataforma adecuada para el proceso de entrenamiento. Para ello, los datos han de ser copiados, lo que afecta al derecho exclusivo de reproducción que ostentan los titulares de las obras.

Es preciso señalar que los datos copiados han de ser almacenados durante cierto periodo de tiempo, lo que dificulta poder sostener que la copia tiene carácter temporal⁶⁹. Para empezar, porque por razones técnicas el conjunto de datos ha de conservarse durante todo el proceso de obtención del modelo de ML⁷⁰. Además, la regulación sobre IA evoluciona en el sentido de exigir cada vez más transparencia a los desarrolladores en relación con los datos de entrenamiento y evaluación utilizados para los modelos de ML⁷¹, lo que implicará que se tengan que almacenar los datos utilizados aún después de haber obtenido el modelo de ML para satisfacer esos requisitos de transparencia.

Ha de tenerse en cuenta, adicionalmente, que si los materiales utilizados para el entrenamiento de un modelo de ML forman parte de una base de datos, se podrían infringir también los derechos exclusivos de extracción y reutilización en aquellas jurisdicciones en que está reconocida la protección a los fabricantes de bases de datos.

En la fase de preparación de datos puede ser necesario también, como se ha visto en el apartado II.3, realizar ajustes o modificaciones en los datos. Como consecuencia, el derecho exclusivo de transformación de los titulares de las obras podría verse afectado.

Como también se ha apuntado, una vez que se obtiene un modelo de ML, este es completamente utilizable sin los datos de entrenamiento y sin los datos de evaluación⁷². Ello implica que el derecho exclusivo de comunicación pública

⁶⁹ En el ámbito de la Unión Europea, esta circunstancia impediría en la generalidad de los casos aplicar la excepción relativa a los actos de reproducción provisional establecida en el artículo 5, apartado 1, de la Directiva 2001/29/CE.

⁷⁰ Como se ha expuesto, no todos los datos recopilados son utilizados en la fase de entrenamiento del modelo, sino que una parte de esos datos (los datos de evaluación) se emplean en la posterior fase de evaluación del modelo, lo que implica que han de conservarse hasta entonces. Se da la circunstancia, además, de que durante la fase de ajuste de hiperparámetros puede optarse, entre otros cambios, por modificar el número de iteraciones en los datos de entrenamiento (esto es, el número de veces que los datos de entrenamiento se «muestran» al modelo). Esto implica la necesidad de conservar el conjunto de datos durante todo el proceso de desarrollo del modelo de ML.

⁷¹ *Vid.*, por ejemplo, el artículo 64.1 de la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión: «Se concederá a las autoridades de vigilancia del mercado acceso a datos y documentación en el contexto de sus actividades, así como pleno acceso a los conjuntos de datos de entrenamiento, validación y prueba utilizados por el proveedor, incluso mediante interfaces de programación de aplicaciones («API», por sus siglas en inglés) u otros medios técnicos y herramientas adecuados que permitan el acceso a distancia».

⁷² J. DREXL *et ál.*, «Technical Aspects of Artificial Intelligence...», cit., p. 8.

de los titulares de las obras utilizadas para el desarrollo del modelo no debería verse, en principio, afectado.

En atención a lo expuesto hasta ahora, todo apunta a que la utilización de materiales protegidos por DPI para el desarrollo de sistemas de ML afecta, en principio, a facultades exclusivas de los titulares. Sin perjuicio de lo anterior, se ha planteado la cuestión de si cabe considerar que el entrenamiento de sistemas de ML no es una actividad relevante para la propiedad intelectual en atención al uso que se hace de las obras y prestaciones, siguiendo el argumento que ahora se expone.

Un principio fundamental de la propiedad intelectual es que las ideas o hechos, como tales, no se protegen mediante DPI, sino que lo que se protege es, únicamente, la expresión de esas ideas o hechos⁷³. A este principio se le denomina la «dicotomía idea-expresión» o la «dicotomía hecho-expresión». Esta distinción entre las ideas o hechos y su expresión es relevante a efectos del empleo de materiales protegidos por DPI para el entrenamiento de sistemas de ML, en la medida en que podría argumentarse que los sistemas de ML están generalmente «interesados» en los hechos o en la información incluidos en las obras y prestaciones que forman parte del conjunto de datos de entrenamiento y no en su expresión. Por ello, del mismo modo que se permite que cualquier persona lea un texto como medio para acceder a los hechos o a la información sin que el acto específico de leer requiera el consentimiento del autor, podría sostenerse que debería permitirse la copia por parte de un sistema de ML en la medida en que esa copia se hace con la finalidad de acceder a los hechos o a la información, esto es, al contenido no protegido por la propiedad intelectual.

Por ilustrar lo anterior con un ejemplo, pensemos en el desarrollo de un sistema de visión artificial que debe aprender a reconocer imágenes de semáforos, para ser posteriormente incorporado a un vehículo de conducción autónoma. El sistema de ML será entrenado con miles de fotografías de semáforos para

⁷³ Artículo 9.2 del Acuerdo sobre los Aspectos de los Derechos de Propiedad Intelectual relacionados con el Comercio (ADPIC) y artículo 2 del Tratado de la OMPI sobre Derecho de Autor (TODA): «*La protección del derecho de autor abarcará las expresiones pero no las ideas, procedimientos, métodos de operación o conceptos matemáticos en sí*». Artículo 2.8 del Convenio de Berna: «*La protección del presente convenio no se aplicará a las noticias del día ni de los sucesos que tengan el carácter de simples informaciones de prensa*». La razón de ser de esta distinción la explica bien Bercovitz: «*La propiedad intelectual no protege las ideas, ni la información (el art. 2.8 del Convenio de Berna excluye expresamente de su ámbito de aplicación las informaciones y las noticias), cuya libre disponibilidad es esencial para el desarrollo del conocimiento social, cultural, económico y científico: no se puede permitir que el primero que descubra una idea o consiga una información detente frente a los demás monopolio o derecho de exclusiva, por mucha importancia y utilidad que tenga ese descubrimiento (esa idea o información), o por mucho reconocimiento social que merezca*». R. BERCOVITZ RODRÍGUEZ-CANO, «*La obra*», *Manual de Propiedad Intelectual* (Coord. R. Bercovitz Rodríguez-Cano), 9^a edición, Tirant Lo Blanch, Valencia, 2019, p. 58.

que aprenda a identificar patrones que le permitan predecir cuándo una imagen corresponde a un semáforo. Estas imágenes deberán ser copiadas y, en su caso, transformadas como parte del proceso de entrenamiento del algoritmo. Sin embargo, en este supuesto queda claro que el algoritmo no está «interesado» en la expresión (es decir, en las características artísticas que pueda, en su caso, tener cada fotografía incluida en el conjunto de datos), sino en acceder al concepto que subyace tras la expresión (es decir, qué hace que un semáforo sea un semáforo). Con base en este razonamiento, se puede argumentar que el ML debería estar fuera del alcance de las facultades exclusivas otorgadas a los titulares de DPI.

Esta ha sido la posición de voces muy autorizadas de la doctrina⁷⁴ en relación con una tecnología denominada minería de textos y datos («MTD»), que consiste —como se verá con mayor detalle en la sección siguiente— en la extracción de conocimiento a partir del análisis de conjuntos de datos y que planteaba, al igual que el ML, el inconveniente de que el sistema solo podía «leer» las obras para acceder a los conceptos subyacentes realizando una copia. Esta posición consistente en dejar la minería de textos y datos fuera del alcance de la propiedad intelectual fue sintetizada en la doctrina mediante la conocida máxima «*el derecho a leer debería ser el derecho a minar*» (*the right to read should be the right to mine*⁷⁵).

Se pueden formular, sin embargo, dos objeciones a la conclusión anterior en relación con el desarrollo de sistemas de ML. La primera, es que incluso si el propósito de copiar una obra protegida por DPI es acceder a la parte no protegida de la obra, esa copia no consentida por el titular infringiría el derecho exclusivo de reproducción (e idéntico razonamiento cabría realizar en relación con las otras facultades exclusivas que se puedan ver afectadas en la fase de entrenamiento)⁷⁶. Este es un argumento que también se podía efectuar en relación con la MTD pero que, sin embargo, no impidió que primara entre la doctrina la consideración de que la MTD debía quedar fuera del alcance de la propiedad intelectual.

⁷⁴ R. HILTY y H. RICHTER, «Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules Part B Exceptions and Limitations (Art. 3 — Text and Data Mining)», *MPI Innovation and Competition Research Paper No. 17-02*, 2019, p.3. T. MARGONI, M. KRETSCHMER, «A deeper look into the EU Text and Data Mining exceptions: Harmonisations, data ownership, and the future of technology», *CREATe Working Paper 2021/7*, pp. 21-22.

⁷⁵ P. MURRAY-RUST, *The Right to Read is the Right to Mine*, 1 junio 2012, <https://blog.okfn.org/2012/06/01/the-right-to-read-is-the-right-to-mine/>

⁷⁶ Como recuerda Margoni, normalmente cualquier copia tiene el potencial de infringir los derechos de autor y, de hecho, la copia temporal de una página web hecha en la memoria caché de ordenadores y tabletas sólo es posible, en el ámbito de la Unión Europea, porque existe una excepción que así lo autoriza—. T. MARGONI, «Artificial Intelligence, Machine Learning and EU copyright law: Who owns AI?», *CREATe Working Paper 2018/12*, diciembre 2018, p. 3.

La segunda objeción que se puede formular a la tesis de considerar el entrenamiento de sistemas de ML fuera del alcance de la propiedad intelectual es específica del ML, y no ha sido abordada de forma tan clara por la doctrina. En el ML ocurre que, en muchos casos, la copia no tiene como finalidad acceder a los elementos no protegidos de la obra (los hechos o la información), sino precisamente a los elementos expresivos. Un buen ejemplo es el proyecto «El Nuevo Rembrandt», explicado al inicio de este trabajo, en el que se buscaba obtener un sistema de IA que pintara con el estilo de Rembrandt. Para ello, el algoritmo fue entrenado con pinturas del artista para que aprendiera a reconocer qué era lo que hacía que un cuadro de Rembrandt fuera tal: la forma de pintar los rasgos faciales, el manejo de luces y sombras, etc. En el proyecto «El Nuevo Rembrandt» el desarrollo del sistema de ML afectó a la parte de las obras de Rembrandt que sí estaría protegida por la propiedad intelectual (la expresión), por lo que difícilmente se puede sostener que en todos los casos de ML la copia es una actividad irrelevante desde la perspectiva de la propiedad intelectual.

En la medida en que ciertas formas de ML sí que afectan a elementos protegidos por la propiedad intelectual, es razonable concluir que el entrenamiento de sistemas de ML es una actividad comprendida en el alcance de la propiedad intelectual, lo que implica que el uso de materiales protegidos por DPI con esa finalidad sin la autorización de los titulares puede constituir una infracción si no existe una excepción o limitación a los DPI que ampare ese uso. Esta ha sido la visión del legislador en varias jurisdicciones, donde partiendo de que el aprendizaje de sistemas de IA es relevante para la propiedad intelectual, se han regulado excepciones o limitaciones a los DPI en relación con esta cuestión, como a continuación se verá.

IV. ESTUDIO DE LAS DISTINTAS APROXIMACIONES A LA CUESTIÓN: UNIÓN EUROPEA, ESTADOS UNIDOS Y JAPÓN

1. APUNTES PRELIMINARES

El uso de materiales protegidos por DPI con el propósito de desarrollar sistemas de IA sin la autorización de sus titulares ha generado un intenso debate. Las jurisdicciones en las que se ha planteado la necesidad de adaptar la legislación sobre propiedad intelectual para abordar esta cuestión de manera específica (principalmente, aquellas en las que existe un mayor desarrollo de la IA) se han encontrado ante el dilema de cómo alcanzar un equilibrio adecuado entre el interés público en favorecer la innovación, de un lado, y el respeto a los DPI como incentivo para la creación literaria, artística y científica, de otro lado.

Importa destacar que el análisis de la cuestión que nos ocupa en el plano internacional no se ha realizado hasta ahora generalmente desde la perspectiva del ML, sino desde la perspectiva de otra tecnología: la minería de textos y datos

(«MTD»). La razón de ello es que la MTD se desarrolló antes que el ML⁷⁷, lo que explica que el debate sobre la utilización de obras protegidas por propiedad intelectual para actividades de MTD tuviera lugar con anterioridad al debate sobre ML que nos ocupa.

La MTD es una rama de la minería de datos (*data mining*), que nace de la intersección entre la estadística y la IA (en particular, el ML). La MTD y el ML tienen en común que ambos permiten extraer valor a partir del análisis de los datos y que ambos permiten obtener buenos resultados en el reconocimiento de patrones. Sin embargo, se trata de técnicas diferentes, donde varía notablemente el modo en que se «enseña» al programa a reconocer esos patrones: la MTD es una técnica donde se precisa de mayor intervención humana, pues el sistema carece de capacidad de autoaprendizaje y sigue un conjunto predefinido de reglas y condiciones para resolver un problema⁷⁸. Además, como técnica, la MTD implica generalmente el uso de ML, pero ello no ocurre así en sentido inverso: el ML puede utilizar técnicas de MTD, pero no necesariamente tiene que hacerlo⁷⁹.

La MTD y el ML también difieren en su finalidad. En la MTD «*los investigadores usan algoritmos para «leer» un gran número de textos (por ejemplo, trabajos científicos sobre química) y extraer automáticamente conocimientos (ello permite, por ejemplo, hallar datos que no se mencionan explícitamente en ninguno de los trabajos, pero que pueden deducirse de todo el corpus)*⁸⁰. Se busca, pues, la extracción de conocimientos a partir de un volumen elevado de datos: encontrar tendencias, clasificaciones, asociaciones o anomalías en los datos que, sin la aplicación de estas técnicas, sería difícil descubrir. El propósito del ML es generalmente distinto: se persigue generar un algoritmo que, tras ser entrenado con un conjunto de datos, sea capaz de aplicarse a nuevos datos para realizar una determinada tarea (por ejemplo, pintar como Rembrandt, componer canciones, reconocer correos de *spam* o distinguir imágenes).

En atención a esa distinta finalidad explicada, los retos que plantean la MTD y el ML para la propiedad intelectual no son exactamente los mismos. En la MTD es habitual que la copia de los materiales protegidos por DPI se realice como

⁷⁷ El ML, como se ha visto en la sección II anterior, aceleró su evolución a partir de 2010 con los avances en la rama del *deep learning* o aprendizaje profundo y a partir de 2015 vivió un auténtico *boom*.

⁷⁸ J. RAMZAI, *Clearly Explained: How Machine learning is different from Data Mining*, 3 mayo 2020. <https://towardsdatascience.com/clearly-explained-how-machine-learning-is-different-from-data-mining-4ee0e0c91bd4> S. GUPTA, *Data Mining vs Machine Learning: What is the Difference?*, 25 febrero 2020. <https://in.springboard.com/blog/data-mining-vs-machine-learning/>

⁷⁹ J. RAMZAI, *Clearly Explained: How Machine learning...*, op. cit., S. GUPTA, *Data Mining vs Machine Learning...*, op. cit.

⁸⁰ Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones sobre Inteligencia artificial para Europa, Bruselas, 25 de abril de 2018, COM(2018) 237 final, p. 10.

medio para acceder al contenido no protegido por la propiedad intelectual (la información o los hechos). En el ML puede darse también esta circunstancia (por ejemplo, en el caso de un sistema de reconocimiento de imágenes que copie fotografías de semáforos para acceder al concepto subyacente y aprender así a distinguirlos), pero puede darse también el caso de que la copia persiga acceder al contenido que estaría protegido por DPI (así ocurría, por ejemplo, en el proyecto «El Nuevo Rembrandt»). Además, como resultado del ML se obtiene, como se ha visto, un algoritmo que es capaz de crear contenidos con un cierto grado de autonomía, lo que puede afectar al mercado potencial de las obras que se han utilizado para el entrenamiento del algoritmo o al propio valor de dichas obras en el mercado.

La conclusión de lo anterior es que, en el debate sobre la cuestión, es preciso atender al concreto uso que se va a realizar de los materiales protegidos por DPI, sin que sea posible equiparar, sin mayores consideraciones, la utilización de esos materiales para actividades de MTD al uso de esos mismos materiales para entrenar sistemas de ML⁸¹. Y ello por cuanto, como se ha visto, existen algunos supuestos de ML que tienen un potencial lesivo sobre los intereses de los titulares de los materiales protegidos por la propiedad intelectual que se utilizan en el entrenamiento mayor que el que pueda, en su caso, presentar la MTD, y que pueden no haber sido abordados de manera específica por la legislación, la jurisprudencia o la doctrina, al representar avances más recientes.

Sin perjuicio de lo anterior, los desarrollos legislativos, jurisprudenciales y doctrinales habidos hasta la fecha en materia de MTD sin duda son relevantes para estudiar la cuestión debatida. Primero, porque en aquellas jurisdicciones donde no se haya tratado específicamente el uso de materiales protegidos por DPI para el ML, habrá que partir de la legislación existente en materia de MTD, en su caso, para estudiar el posible encaje del ML⁸². Y segundo, porque las soluciones legislativas y jurisprudenciales propuestas

⁸¹ Es importante señalar, sin embargo, que algunas definiciones de MTD contenidas en textos legislativos son tan amplias que en ellas cabe incluir, prácticamente, cualquier método de análisis computacional de datos (entendiéndose comprendidas también, por ende, las técnicas de ML). Por tanto, aunque en este artículo se han expuesto las diferencias entre la MTD y el ML como tecnologías distintas, aunque relacionadas en algunos supuestos (como se ha expuesto, la MTD suele implicar la utilización de ML, si bien el ML no implica siempre emplear MTD), habrá que atender en cada caso y para evitar confusiones terminológicas, a la acepción de MTD empleada. En todo caso, coincidimos con Carroll en que, para no incurrir en estas confusiones en la terminología empleada, sería conveniente evitar el uso de la expresión «minería de textos y datos» y, en su lugar, utilizar expresiones más amplias como «investigación computacional» o «minería de contenido», en las que quepan todo tipo de técnicas. M. W. CARROLL, «Copyright and the Progress of Science: Why Text and Data Mining is Lawful», *Washington College of Law Research Paper No. 2020-15*, p. 895. En nuestra opinión, la denominación «análisis computacional de datos» sería también una buena alternativa, y garantizaría que todos los futuros desarrollos tecnológicos en el ámbito de la IA quedaran comprendidos en la definición.

⁸² *Vid.* nota 81 *supra*.

hasta la fecha en materia de MTD probablemente apuntarán el camino hacia el que evolucionará el tratamiento de materiales protegidos por DPI para ML en esas jurisdicciones.

En las siguientes páginas se estudiarán los casos de la Unión Europea, Estados Unidos y Japón⁸³. Como se verá, las soluciones alcanzadas no han sido homogéneas. Así, mientras que algunas jurisdicciones han resuelto la tensión entre la propiedad intelectual y la innovación apostando por la segunda y permitiendo, con carácter general, la utilización de obras protegidas por DPI sin autorización de los titulares (es el caso de Japón o Estados Unidos), otras han optado por soluciones más volcadas en la protección de los DPI que, inevitablemente, suponen cierto freno a la innovación (es el caso de la Unión Europea).

El análisis comparado resulta particularmente útil para resolver una cuestión compleja como la que aquí se plantea, a fin de explorar distintas posibilidades de solución para tratar de encontrar la más adecuada en atención a los intereses cuya protección se quiera primar. Además, ha de tenerse en cuenta que las decisiones que tomen los legisladores de cada una de las economías desarrolladas en la materia estudiada pueden verse significativamente influenciadas por las que tomen el resto, por los motivos explicados en la introducción de este trabajo. Por eso, no consideramos que la materia estudiada sea, ni mucho menos, una cuestión que pueda darse por cerrada.

A nadie escapa que, en la carrera actual hacia el liderazgo en IA, el hecho de que un país se presente como especialmente favorable para la innovación puede resultar en la atracción de talento e inversión. En particular, el hecho de que una empresa que desarrolle sistemas de ML tenga que pagar o no licencia para utilizar materiales protegidos por DPI es un factor que puede condicionar la viabilidad económica del proyecto —principalmente en el caso de que la empresa sea de tamaño pequeño o mediano, como suele ser el caso de las *start-up*— y provocar su fuga hacia una jurisdicción que le permita evitar ese coste. Si esa huida de talento y emprendimiento es generalizada, no sería descartable que se revisara la legislación de propiedad intelectual del país que padece la fuga de cerebros a fin de retener el talento (o, incluso, atraerlo) y fomentar así el impulso a la innovación. Este escenario no es descartable en el caso de la Unión Europea, por las razones que ahora se expondrán.

⁸³ Existen otras jurisdicciones que han legislado en esta materia o que están revisando su legislación de propiedad intelectual en la actualidad. Ante la imposibilidad de abordar en este artículo todas las regulaciones, se ha considerado que los ejemplos de la Unión Europea, Estados Unidos y Japón resultan de interés no sólo por su relevancia política, económica y tecnológica, sino porque representan tres maneras muy distintas de abordar la cuestión, que sin duda inspirarán futuros desarrollos legislativos en otros ordenamientos.

2. UNIÓN EUROPEA

La Unión Europea, en línea con su objetivo de conseguir un mercado único digital para Europa⁸⁴, se ocupó de dar respuesta legislativa⁸⁵ a la tensión entre la minería de textos y datos (MTD) y la propiedad intelectual en la Directiva 2019/790, del Parlamento Europeo y del Consejo, de 17 de abril de 2019, sobre los derechos de autor y derechos afines en el mercado único digital («DDAMUD»).

Como se explica en el considerando 8 de la DDAMUD, la iniciativa legislativa en este ámbito partía del reconocimiento de que la MTD podía favorecer la actividad de la comunidad investigadora y suponer un apoyo a la innovación, pero los organismos e instituciones de investigación se enfrentaban a cierta inseguridad jurídica en relación con los contenidos que podían utilizar para llevar a cabo actividades de MTD —particularmente, cuando estaban protegidos por DPI—. Esta inseguridad jurídica se explicaba porque, si bien con anterioridad a la DDAMUD las actividades de MTD con fines de investigación científica podían considerarse cubiertas por determinadas excepciones y limitaciones de los DPI, éstas tenían carácter optativo y no estaban plenamente adaptadas a la investigación científica (considerando 10)⁸⁶.

La falta de claridad sobre el encaje de las actividades de MTD en el régimen de excepciones y limitaciones de los DPI con fines de investigación científica, junto con la importante fragmentación normativa existente⁸⁷, suponía un fre-

⁸⁴ En el documento «Una Estrategia para el Mercado Único Digital de Europa» se identificaba la necesidad de adoptar «*un marco europeo moderno de los derechos de propiedad intelectual*» y se explicaba que «*la innovación en la investigación tanto para fines comerciales como no comerciales, basada en la utilización de la extracción de textos y datos (por ejemplo, copiando textos y conjuntos de datos para buscar correlaciones o acontecimientos significativos) puede verse obstaculizada por un marco jurídico confuso y enfoques divergentes a escala nacional*». Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones, Una Estrategia para el Mercado Único Digital de Europa, COM (2015) 192 final, Bruselas, 6 de mayo de 2015, pp. 7-8.

⁸⁵ Toda vez que, en el ámbito de la UE, existe legislación específica sobre la materia analizada, no se considera imprescindible a efectos de este trabajo el estudio de algunos casos resueltos por el TJUE en los que puedan tratarse cuestiones relacionadas con las aquí tratadas, como *Infopaq International A/S v. Danske Dagblades Forening* (asunto C-5/08), *Innoweb BV v. Wegener ICT y Wegener Mediaventions BV* (asunto C-2020/12) o *CV-Online Latvia SIA v. Melons SIA*.

⁸⁶ Para un estudio del régimen de excepciones y limitaciones que podían resultar de aplicación a las actividades de MTD en el Derecho de la Unión con anterioridad a la DDAMUD, *vid. C. GEIGER, G. FROSIO y O. BULAYENKO, «Text and Data Mining: Articles 3 y 4 of the Directive 2019/790/EU», Propiedad intelectual y mercado único digital europeo* (Dir. C. Saiz García y R. Evangelio Llorca), Tirant Lo Blanch, Valencia, 2019, pp. 36-56. También sobre esta cuestión, *vid. A. GARCÍA VIDAL, «Propiedad intelectual y minería de textos: Estudio de los artículos 3 y 4 de la Directiva (UE) 2019/790», Actas de Derecho Industrial 40, 2019-2020*, pp. 99-124.

⁸⁷ Importa señalar que algunos Estados miembros habían ya adoptado, a nivel nacional, normas que establecían excepciones y limitaciones para actividades de MTD. Es el caso de Reino Unido (en 2014), Francia (en 2016), Estonia (en 2017) y Alemania (en 2017). Para un estudio más de-

no para la consecución de un mercado único digital a nivel europeo. Además, como señalan Geiger, Frosio y Bulayenko, existía una gran diversidad en las prácticas de concesión de licencias que conllevaba la asunción de elevados costes de transacción⁸⁸.

Para remediar esta situación, la DDAMUD introdujo dos excepciones específicas para actividades de MTD en sus artículos 3 y 4, que se estudiarán a continuación⁸⁹. Conviene avanzar que, si bien la doctrina acogió de buen grado la iniciativa del legislador europeo de regular esta materia de forma armonizada, la concreta solución adoptada recibió la crítica de un importante sector por considerar que conllevaba un freno a la innovación.

La DDAMUD define la MTD en su artículo 2.2 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*». Como apuntan acertadamente Margoni y Kretschmer, esta definición de MTD es tan amplia que permite abarcar la mayor parte de actividades donde las tecnologías digitales se utilicen para analizar información, incluyendo el ML⁹⁰.

Las excepciones establecidas en los artículos 3 y 4 de la DDAMUD tienen carácter obligatorio y comprenden tanto el derecho exclusivo de reproducción de los titulares de DPI como el derecho *sui generis* de extracción de bases de datos, así como el derecho reconocido a las editoriales de publicaciones de prensa en el primer apartado del artículo 15 de la DDAMUD. En la aplicación de estas excepciones resulta de aplicación la «regla de los tres pasos»⁹¹, así como determinados apartados del artículo 6, apartado 4, de la Directiva 2001/29/CE, mediante el que se pretende que los beneficiarios de excepciones y limitaciones no vean impedida esta posibilidad por la

tallado sobre estas regulaciones, *vid.* C. GEIGER, G. FROSIO y O. BULAYENKO, «Text and Data Mining... », op. cit., pp. 50-53.

⁸⁸ C. GEIGER, G. FROSIO y O. BULAYENKO, «Text and Data Mining... », op. cit., p. 45.

⁸⁹ En el procedimiento legislativo de las excepciones de MTD tuvieron lugar diversas modificaciones que ponen de manifiesto el intenso debate que se generó en torno a esta cuestión. Para una exposición más detallada al respecto, *vid.* B. GONZÁLEZ OTERO, «Las excepciones de minería de textos y datos más allá de los derechos de autor: la ordenación privada contraataca», *Propiedad intelectual y mercado único digital europeo* (Dir. C. Saiz García y R. Evangelio Llorca), Tirant Lo Blanch, Valencia, 2019, pp. 83-89.

⁹⁰ No obstante, como señalan estos autores, las excepciones de los artículos 3 y 4 DDAMUD plantean problemas prácticos en su aplicación al ML. *Vid.* T. MARGONI, M. KRETSCHMER, «A deeper look into the EU Text and Data Mining exceptions: Harmonisations, data ownership, and the future of technology», CREATe Working Paper 2021/7, pp. 6-7, 24-28.

⁹¹ Así lo establece el artículo 7, apartado 2 de la DDAMUD, al remitirse al artículo 5, apartado 5 de la Directiva 2001/29/CE. Conforme a esta regla, las excepciones y limitaciones únicamente se aplicarán (i) en determinados casos concretos, (ii) que no entren en conflicto con la explotación normal de la obra o prestación y (iii) que no perjudiquen injustificadamente los intereses legítimos del titular del derecho.

adopción por parte de los titulares de los derechos de medidas tecnológicas de protección⁹².

Ha de tenerse en cuenta, además, que las excepciones de los artículos 3 y 4 parten del presupuesto común de que las entidades que desarrollen actividades de MTD tendrán acceso lícito⁹³ a los contenidos.

El artículo 3, en primer lugar, introduce una excepción que resulta de aplicación a las actividades de reproducción y extracción llevadas a cabo por «organismos de investigación e instituciones del patrimonio cultural con el fin de realizar, con fines de investigación científica, minería de textos y datos de obras u otras prestaciones a las que tengan acceso lícito». Nótese, pues, que para que sea aplicable la excepción ha de cumplirse un doble requisito: que las actividades de MTD sean realizadas por «organismos de investigación»⁹⁴ o «instituciones del patrimonio cultural»⁹⁵ (requisito de carácter subjetivo) y que se realicen con «fines de investigación científica» (requisito de carácter objetivo).

Es relevante señalar que la excepción para actividades de MTD contemplada en el artículo 3 de la DDAMUD no puede excluirse por acuerdo entre las partes. Cualquier disposición contractual entre el titular de derechos y las entidades comprendidas en el ámbito subjetivo de aplicación del artículo 3 por la que se

⁹² Así lo prevé el artículo 7, apartado 2 de la DDAMUD, al remitirse a los párrafos primero, tercero y quinto del artículo 6, apartado 4, de la Directiva 2001/29. *Vid.* B. GONZÁLEZ OTERO, «Las excepciones de minería de textos y datos...», op. cit., p. 18-20.

⁹³ Así lo explica el considerando 14 en relación con los organismos de investigación e instituciones responsables del patrimonio cultural: «debe entenderse el acceso lícito como el acceso a contenidos basado en una política de acceso abierto o por medio de disposiciones contractuales entre titulares de derechos y organismos de investigación o instituciones responsables del patrimonio cultural, como suscripciones, o por otros medios lícitos. Por ejemplo, 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. El acceso lícito también comprende el acceso a contenidos disponibles de forma gratuita en línea». El considerando 18 se refiere también a la necesidad de licitud en el acceso a los contenidos al referirse a las actividades de MTD desarrolladas por el resto de entidades: «Esta excepción o limitación solo debe aplicarse cuando el beneficiario acceda de forma legítima a la obra u otra prestación, también cuando se haya puesto a disposición del público en línea, y en la medida en que los titulares de derechos no se hayan reservado de forma adecuada los derechos de hacer reproducciones y extracciones con fines de minería de textos y datos».

⁹⁴ La definición de «organismo de investigación» a efectos de la DDAMUD se incluye en el apartado 1 del artículo 2: «una universidad, incluidas sus bibliotecas, un instituto de investigación o cualquier otra entidad cuyo principal objetivo sea realizar investigaciones científicas o llevar a cabo actividades educativas que también impliquen realizar investigaciones científicas».

a) sin ánimo de lucro o reinvirtiendo todos los beneficios en sus investigaciones científicas, o
b) conforme a una misión de interés público reconocida por un Estado miembro, de tal manera que una empresa que ejerza una influencia decisiva en dicho organismo no pueda gozar de acceso preferente a los resultados generados por tales investigaciones científicas».

⁹⁵ La definición de «institución responsable del patrimonio cultural» a efectos de la DDAMUD se incluye en el apartado 3 del artículo 2: «una biblioteca o un museo accesibles al público, un archivo o una institución responsable del patrimonio cinematográfico o sonoro».

pretenda modificar la excepción será inaplicable y carecerá de efectos (artículo 7, apartado 1, DDAMUD).

Se permite a los titulares de derechos la aplicación de medidas tecnológicas de protección «para garantizar la seguridad e integridad de las redes y bases de datos en que estén almacenadas las obras u otras prestaciones», si bien se señala que «dichas medidas no irán más allá de lo necesario para lograr ese objetivo» (artículo 3, apartado 3).

La excepción del artículo 3 contempla, por último, que las copias de obras y otras prestaciones efectuadas puedan conservarse con fines de investigación científica y, en particular, para la verificación de resultados de la investigación.

Junto a la excepción del artículo 3 —que tiene, como se ha visto, un ámbito de aplicación muy restringido por los sujetos a quienes se aplica y por los fines que ha de tener la MTD—, la DDAMUD establece en el artículo 4 una excepción de carácter más general para las actividades de MTD, aplicable en aquellos casos en que no lo sea la excepción del artículo 3 —esencialmente, cuando la MTD se realice por entidades distintas a las allí contempladas, lo que incluye las entidades privadas, o se realice con fines comerciales—.

La excepción del artículo 4 de la DDAMUD presenta una diferencia esencial con respecto a la del artículo 3: se aplica solo a condición de que «el uso de las obras y otras prestaciones a que se refiere dicho apartado no esté reservado expresamente por los titulares de derechos de manera adecuada⁹⁶, como medios de lectura mecánica en el caso del contenido puesto a la disposición del público en línea». En otras palabras: la excepción del artículo 4 queda en manos de los titulares de las obras, que pueden aplicar medidas tecnológicas de protección y excluir por vía contractual la aplicación de la excepción (mecanismo de *opt-out*).

Adicionalmente, y en contraste con lo establecido en el artículo 3 —que permite, como se ha visto, conservar las copias de obras y prestaciones con fines de investigación científica—, el artículo 4 permite la conservación de las reproducciones y extracciones solamente durante el tiempo que sea necesario para realizar las actividades de MTD.

La regulación de las excepciones para MTD de la DDAMUD recibió críticas por parte de un sector importante de la doctrina europea, alineada mayorita-

⁹⁶ Como apunta Geiger, no está claro en el texto de la Directiva si las herramientas «adecuadas» para ejercer el *opt-out* son las tecnológicas, las contractuales o ambas. C. GEIGER, «The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and The Failed Text and Data Mining Limitations in the CSDM Directive», *American University Washington College of Law*, PIJIP/TLS Research Paper, 66, abril 2021, p.9.

riamente en torno a la tesis antes expuesta de que «el derecho a leer debería ser el derecho a minar». Esto es, de considerar que la MTD debería ser un uso siempre permitido de las obras protegidas por DPI, del mismo modo que lo es la lectura de esas mismas obras por parte de un ser humano con el fin de extraer ideas o información⁹⁷. Se apoyó, incluso, el reconocimiento de un derecho a la lectura de las máquinas⁹⁸.

No fue esa, como se ha visto, la postura del legislador europeo que, partiendo de la premisa de que el derecho a leer *no* es el derecho a minar y de que, por ende, las actividades de MTD sí que son relevantes para la propiedad intelectual, diseñó un sistema de excepciones que, por su reducido ámbito, supone un obstáculo a la innovación⁹⁹. El restringido alcance de la excepción del artículo 3 conlleva que una parte muy relevante de las actividades de MTD (las desarrolladas por entidades privadas con fines comerciales) queden cubiertas únicamente por la excepción del artículo 4, que permite que los titulares de las obras puedan dejar sin efecto la excepción contractualmente o mediante medidas tecnológicas de protección¹⁰⁰. En relación con estas medidas, se ha criticado también que la DDAMUD no haya prohibido expresamente que las excepciones de MTD queden invalidadas mediante medidas tecnológicas de protección, en lugar de hacer una remisión al artículo 6, apartado 4, de la Directiva 2001/29, que no aporta seguridad jurídica a quienes quieran hacer uso de la excepción¹⁰¹.

Adicionalmente, apunta González Otero que la Directiva podría también haber clarificado que las excepciones de MTD de la DDAMUD son también de aplicación a los materiales que, aun no estando protegidos por DPI, pueden ser objeto de reserva por parte de sus titulares por vía contractual o por medio de medidas tecnológicas de protección, a fin de evitar lo ocurrido en la sentencia del TJUE del caso *Ryanair*¹⁰². En esta sentencia¹⁰³, el Tribunal

⁹⁷ R. HILTY y H. RICHTER, «Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules Part B Exceptions and Limitations...», op. cit., p.5; C. GEIGER, G. FROSIO, O. BULAYENKO, «The exception for Text and Data Mining (TDM) in the proposed Directive on copyright in the digital single market – legal aspects», *Center for International Intellectual Property Studies Research Paper No. 2018-02*, p. 32; R. DUCATO, A. STROWEL, «Limitations To Text and Data Mining and Consumer Empowerment. Making the Case for a Right to «Machine Legibility»», *CRIDES Working Paper Series*, 31 de octubre de 2018, pp. 20-21.

⁹⁸ R. DUCATO, A. STROWEL, «Limitations To Text and Data Mining... », op. cit., pp. 28-33.

⁹⁹ Así lo explica Rosati, al reflexionar sobre si la propiedad intelectual debe ser un obstáculo o un facilitador para la innovación al hilo de las excepciones para MTD de la DDAMUD. E. ROSATI, «Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity», *Asia Pacific Law Review*, 26 de septiembre de 2019.

¹⁰⁰ C. GEIGER, G. FROSIO y O. BULAYENKO, «Text and Data Mining... », op. cit., p. 56-61.

¹⁰¹ B. GONZÁLEZ OTERO, «Las excepciones de minería de textos y datos...», op. cit, p. 20.

¹⁰² B. GONZÁLEZ OTERO, «Las excepciones de minería de textos y datos...», op. cit, p. 20.

¹⁰³ Caso Ryanair, TJUE, C-30/14, Ryanair Ltd. V PR Aviation BV (2015), ECLI:EU:C:2015:10.

concluyó que la base de datos de Ryanair no está protegida por el derecho de autor ni por el derecho sui generis de bases de datos, pero que el Derecho de la Unión «*no se opone a que el creador de dicha base de datos establezca limitaciones contractuales a su utilización por terceros, sin perjuicio del Derecho nacional aplicable*». La interpretación general de esta resolución ha sido, como señala la mencionada autora, que «*el usuario de una base de datos tiene de facto más derechos de acceso a las bases de datos protegidas por derechos de autor que a las no sujetas a protección, restringiendo aún más el uso de la minería de textos y datos*»¹⁰⁴.

En conclusión, existe cierto consenso en la doctrina europea sobre la insuficiencia de las excepciones para MTD de la DDAMUD para permitir que materiales protegidos por la propiedad intelectual sean utilizados en el desarrollo de IA¹⁰⁵. Como apunta de forma muy gráfica Geiger, la UE parece un equipo de fútbol al que le falta un delantero para poder ganar la competición por la IA frente a otras jurisdicciones, al estar las posibilidades de innovación frenadas por el actual régimen de excepciones para MTD de la DDAMUD¹⁰⁶. Por ello, se proponen distintas soluciones, como el reconocimiento de un derecho al procesamiento de datos con fines de ML¹⁰⁷ o de un derecho de remuneración en algunos casos en que se pueda considerar adecuado (por ejemplo, cuando la IA produzca una obra derivada en un contexto comercial)¹⁰⁸.

También se ha planteado si sería posible considerar que la utilización de materiales protegidos por DPI para el entrenamiento de sistemas de ML queda cubierta por la excepción relativa a los actos de reproducción provisional establecida en el artículo 5, apartado 1, de la Directiva 2001/29/CE¹⁰⁹. Al respecto cabe indicar que, si bien no sería descartable que en algún supuesto y en atención a las circunstancias del caso pudiera sostenerse que se cumplen los requisitos cumulativamente previstos en el artículo 5, apartado 1, en la generalidad de los casos resultará difícil sostener que la copia tiene carácter temporal en la medida en que los datos copiados habrán de ser almacenados durante cierto periodo de tiempo por las razones expuestas en la sección II.3 de este trabajo, a la que nos remitimos.

¹⁰⁴ B. GONZÁLEZ OTERO, «Las excepciones de minería de textos y datos...», op. cit, p. 20.

¹⁰⁵ J. DREXL, R. M. HILTY, L. DESAUNETTES-BARBERO, J. GLOBOCNIK, B. GONZÁLEZ OTERO, J. HOFFMAN, D. KIM, S. KULHARI, H. RICHTER, S. SCHEUERER, P. R. SLOWINSKI, K. WIEDEMANN, «Artificial Intelligence and Intellectual Property Law Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate», *Max Planck Institute for Innovation and Competition Research Paper No 21-10*, abril 2021, pp. 8-9. C. GEIGER, «The Missing Goal-Scorers in the Artificial Intelligence Team...», op. cit.

¹⁰⁶ C. GEIGER, «The Missing Goal-Scorers...», op. cit., pp. 1-4.

¹⁰⁷ M. KOP, «The Right to Process Data for Machine Learning Purposes in the EU», *Harvard Journal of Law & Technology*, volumen 34, *Digest Spring 2021*, p. 11-18.

¹⁰⁸ C. GEIGER, «The Missing Goal-Scorers...», op. cit., p. 9 nota 33,

¹⁰⁹ Para un estudio sobre la cuestión, *vid.* T. MARGONI, M. KRETSCHMER, «A deeper look...», op. cit.

3. ESTADOS UNIDOS

En Estados Unidos («EEUU»), a diferencia de lo que ocurre en la UE y en otros países, no se ha establecido una excepción específica para MTD ni, en general, para el desarrollo de sistemas de IA, sino que el encaje de estas actividades en la legislación de propiedad intelectual se ha analizado desde la perspectiva de la doctrina del «uso justo» (*fair use*).

El *fair use* es una doctrina que permite el uso sin licencia de obras protegidas por derechos de autor si se cumplen determinadas condiciones. Se regula en el artículo 107 de la Copyright Act, que establece que para determinar si el uso que se hace de una obra con derechos de autor en un caso particular es un uso justo, deben tomarse en consideración cuatro factores: «(1) *el propósito y el carácter del uso, incluyendo si el uso es de naturaleza comercial o si tiene fines educativos sin ánimo de lucro*; (2) *la naturaleza de la obra protegida por el derecho de autor*; (3) *la cantidad y sustancialidad de la parte utilizada en relación con la obra protegida por derecho de autor en su conjunto*; y (4) *el efecto del uso sobre el mercado potencial o sobre el valor de la obra protegida (...)*»¹¹⁰. Los cuatro factores señalados no son cumulativos, sino que se trata de un elenco de elementos que deben ser tomados en consideración de manera conjunta y que exigen un estudio particular de los hechos y circunstancias de cada caso concreto.

El elemento más importante en la aplicación reciente de la doctrina del *fair use* es la consideración de si un determinado uso es o no «transformador», en el marco del análisis del primer factor. Se entiende que una nueva obra es transformadora —y, por ende, que implica un uso justo de la obra original— cuando la obra original «*es usada como materia prima, transformada en la creación de nueva información, nueva estética, nuevos conocimientos y entendimientos*»¹¹¹. Aunque en la consideración de si la obra constituye un uso justo también se tienen en cuenta el resto de factores, como el efecto del uso en el mercado potencial de la obra protegida, el carácter transformador es, hoy por hoy, el elemento con más peso¹¹²: «*cuanto más transformadora sea la nueva obra, menor será la importancia del resto de factores, como el uso con carácter comercial, que pueden pesar en contra de la consideración de fair use*»¹¹³.

¹¹⁰ Traducción propia.

¹¹¹ Traducción propia de la conocida definición de *fair use* del Juez Leval, acogida por el Tribunal Supremo: «*El uso debe ser productivo y debe emplear la materia citada de una manera diferente o para un propósito diferente al original. (...) Si (...) el uso secundario agrega valor al original - si la materia citada se usa como materia prima, transformada en la creación de nueva información, nueva estética, nuevas percepciones y entendimientos: este es el tipo de actividad que la doctrina del uso justo pretende proteger para el enriquecimiento de la sociedad*». P. N. Leval «Toward a Fair Use Standard», 103 *Harvard Law Review*, 1105, 1990, p. 1111.

¹¹² B. L. W. SOBEL, «*Artificial Intelligence's...*», op. cit., p. 50.

¹¹³ Traducción propia. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994). En B. L. W. SOBEL, «*Artificial Intelligence's...*», op. cit., p. 50.

La doctrina del *fair use* ha sido una gran aliada para el desarrollo tecnológico en EEUU en las últimas décadas. Con carácter general, esta doctrina ha permitido hasta ahora a los ordenadores «leer» (es decir, copiar) trabajos protegidos por DPI. Grimmelman, haciéndose eco de esta permisividad en la aplicación del *fair use* por los tribunales para permitir que los robots puedan «leer» (copiar), escribió de manera crítica que «*silenciosamente, de manera invisible, casi por accidente, el derecho de autor ha concluido que la lectura hecha por robots no cuenta. La infracción es solo para humanos; cuando los ordenadores lo hacen, es un uso legítimo (fair use)*»¹¹⁴.

Para explicar cómo los tribunales estadounidenses han abordado la cuestión, nos referiremos a continuación a algunos de los casos más relevantes en la materia.

Comenzaremos con *Sega v. Accolade*¹¹⁵, que dio lugar a una doctrina acuñada como doctrina del uso «no expresivo» (*non-expressive use*), que siguió desarrollándose a partir de entonces y que es de vital importancia para la cuestión que nos ocupa. El término uso «no expresivo» (también denominado uso «no consuntivo») hace referencia a «*cualquier acto de reproducción que no esté destinado a permitir el disfrute, la apreciación o la comprensión humana de la expresión copiada como expresión*»¹¹⁶. En otras palabras, los usos no-expresivos son «*actos de reproducción que no comunican la expresión original del autor al público*»¹¹⁷.

Accolade era una empresa desarrolladora de videojuegos que quería que sus juegos pudieran funcionar en la videoconsola Sega Genesis (conocida como Mega Drive en algunos territorios fuera de EEUU) sin tener una licencia de Sega (cuyo modelo de negocio consistía, en parte, en licenciar el software a desarrolladores de videojuegos). Con el propósito de evitar el pago de la licencia, Accolade aplicó ingeniería inversa a fin de comprender la manera en que la consola funcionaba con los videojuegos. Para ello, tuvo que copiar buena parte del código de Sega, si bien los videojuegos producidos por Accolade sólo contenían pequeños extractos del código de Sega.

El tribunal consideró que el uso era justo, ya que la copia era «*la única forma de acceder a las ideas y elementos funcionales incorporados en un programa informático con derechos de autor*»¹¹⁸. Dicho de otro modo, Accolade no copiaba el código de Sega para acceder a la expresión (que es lo que el derecho de autor protege), sino para acceder a la información que contenía

¹¹⁴ Traducción propia. J. GRIMMELMAN, «Copyright for Literate Robots», *Iowa Law Review*, 101:657, 2016, p. 658.

¹¹⁵ *Sega Enters. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992).

¹¹⁶ Traducción propia. M. SAG, «The New Legal Landscape for Text Mining and Machine Learning», *Journal of the Copyright Society of the USA*, Vol. 66, febrero 2020, p.6.

¹¹⁷ Traducción propia. M. SAG, «The New Legal Landscape», op. cit., p. 18.

¹¹⁸ Traducción propia. *Sega Enters. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992), 1527.

esa expresión. En la medida en que la copia de la expresión era solo un paso intermedio para acceder a los elementos no protegidos de la obra, se consideró que el uso fue justo. *Sega v. Accolade* sentó el precedente de que la copia no autorizada de una obra protegida por derechos de autor, en tanto en cuanto fuera necesaria para realizar un uso no expresivo, se consideraba un uso justo¹¹⁹.

Otro precedente importante en esta evolución fueron los casos *Perfect 10 v. Amazon*¹²⁰ y *Kelly v. Arriba*¹²¹. Ambos casos versaban sobre motores de búsqueda de imágenes en Internet que mostraban a los usuarios, en respuesta a consultas de búsqueda, miniaturas de imágenes sobre las que los demandantes tenían derechos de autor (y que quedaban almacenadas en los servidores de los demandados). Estos casos diferían de *Sega v. Accolade* en que las compañías demandadas sí que habían copiado elementos expresivos de las obras y comunicaban parte de esa expresión al público¹²². A pesar de lo anterior, en ambos asuntos los tribunales consideraron que el uso de las obras protegidas era un uso transformador, en la medida en que la copia no tenía un propósito estético o artístico, sino que su finalidad era facilitar la búsqueda y el acceso a las imágenes: «*aunque una imagen pueda haber sido creada originalmente para cumplir una función de entretenimiento, estética o informativa, un motor de búsqueda transforma la imagen en un puntero que dirige al usuario a una fuente de información*»¹²³. Los tribunales también razonaron que, en la medida en que las miniaturas ofrecidas no tenían una finalidad artística, tampoco reemplazaban a las obras originales (cuarto factor del *fair use*).

Otro precedente que siguió a los anteriores, de enorme relevancia en lo que a la MTD se refiere, es *Authors Guild v. Google, Inc.*¹²⁴, uno de los famosos casos de Google Books. Google Books era un proyecto de Google que consistía en escanear prácticamente todos los libros del mundo, sin el permiso previo de los titulares de los derechos de autor, para crear un catálogo de biblioteca digital en el que a los lectores que buscaban libros se les mostraban fragmentos de las obras protegidas. Google no obtuvo ningún beneficio comercial directo por el Proyecto Google Books —no había anuncios en Google Books y la compañía no recibía remuneración por proporcionar enlaces a los sitios web donde se podían comprar los libros—. El tribunal concluyó que la copia de los libros por parte de Google era un uso legítimo, ya que implicaba un uso «altamente transformador» de las obras originales, en la medida en que Google creó un

¹¹⁹ J. GRIMMELMAN, «Copyright for Literate Robots», op. cit., p. 658, 664. B. L. W. SOBEL, «Artificial Intelligence's...», op. cit., p. 8.

¹²⁰ *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir. 2007).

¹²¹ *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003).

¹²² B. L. W. SOBEL, «Artificial Intelligence's...», op. cit., p. 8.

¹²³ Traducción propia. *Perfect 10*, 508 F.3d, 1165.

¹²⁴ *Authors Guild, Inc. v. Google, Inc.* (*Authors Guild I*), 954 F. Supp. 2d 282 (S.D.N.Y. 2013).

motor de búsqueda¹²⁵. El tribunal también señaló, en relación con el cuarto factor, que «*Google Books no reemplaza ni sustituye a los libros porque no es una herramienta para leer libros*»¹²⁶.

Interesa destacar que en *Authors Guild v. Google, Inc.* el tribunal aplicó el razonamiento sobre el uso «no expresivo» de la obra para delimitar el efecto del uso en el mercado potencial de la obra protegida (el cuarto factor de la doctrina del *fair use*)¹²⁷. Así, el tribunal apreció que Google Books podía perjudicar el valor de mercado de la obra si, por ejemplo, un usuario estaba interesado en obtener información sobre un acontecimiento histórico y, al ver un fragmento de un libro en Google Books donde se proporcionaba esa información, el usuario perdía el interés en comprar el libro en cuestión¹²⁸. Sin embargo, como la propiedad intelectual no protege ni los hechos ni las ideas, el tribunal concluyó que la eventual afectación al mercado potencial de la obra sería irrelevante para el *copyright*.

El caso de *Authors Guild* supuso un importante aval a las técnicas de MTD, respecto de las que se resaltó su potencial para la innovación y el progreso¹²⁹. Se puede afirmar que, desde *Authors Guild*, en Estados Unidos la MTD pasó a considerarse un uso permitido por la legislación de *copyright*¹³⁰.

¹²⁵ Para un interesante estudio sobre cómo se hubiera resuelto el caso de Google Books en el ámbito del Derecho de la UE atendiendo al régimen entonces vigente de excepciones y limitaciones, *vid.* R. SÁNCHEZ ARISTI, «La sentencia de apelación en el caso Google Books y los límites del fair use: una oportunidad para los autores y editores europeos», *La Ley*, nº 8654, 26 de noviembre de 2015. Como señala el autor, la aproximación al problema desde la perspectiva del Derecho europeo hubiera sido sustancialmente diferente, como lo prueba la Sentencia dictada por el TJUE el 16 de julio de 2009 en el caso *Infopaq International A/S v. Danske Dagblades Forening* (asunto C-5/08), que guarda algunas semejanzas con el caso de Google Books.

¹²⁶ Traducción propia. *Authors Guild I*, 954 F. Supp. 2d, 291.

¹²⁷ B. L. W. SOBEL, «Artificial Intelligence's...», op. cit., p. 10.

¹²⁸ B. L. W. SOBEL, «Artificial Intelligence's...», op. cit., p. 10.

¹²⁹ «En segundo lugar, además de ser una importante herramienta de referencia, Google Books promueve en gran medida una tipo de investigación denominada «minería de datos» o «minería de texto». Google Books permite a los académicos de humanidades analizar cantidades masivas de datos: el registro literario creado por un colección de decenas de millones de libros. Los investigadores pueden examinar frecuencias de palabras, patrones sintácticos y marcadores temáticos para estudiar cómo ha cambiado el estilo literario con el tiempo. Usando Google Books, por ejemplo, los investigadores pueden medir la frecuencia de las referencias a los Estados Unidos como una sola entidad («los Estados Unidos es») versus referencias a los Estados Unidos en el plural («los Estados Unidos son») y cómo ese uso ha cambiado con el tiempo. La habilidad para determinar la frecuencia con la que aparecen diferentes palabras o frases en los libros en diferentes épocas puede proporcionar conocimientos en campos tan diversos como la lexicografía, la evolución de la gramática, la memoria colectiva, la adopción de la tecnología, la búsqueda de la fama, la censura y la historia epidemiológica». Traducción propia. *Authors Guild, Inc. v. Google Inc.*, 954 F. Supp. 2d 282 (N.Y.S.D. 2013), at 288.

¹³⁰ M. SAG, «The New Legal Landscape...», op. cit., p. 23.

La reciente sentencia dictada en el caso *Oracle v. Google*¹³¹ supone otro precedente más en la línea favorable de pronunciamientos de los tribunales estadounidenses en pos de la innovación y la interpretación amplia del uso transformador. En este caso se discutía si la utilización por parte de Google, en el desarrollo del sistema operativo Android, de las APIs (*Application Programming Interface*) de Java sin el consentimiento de su titular (Oracle) suponía una infracción de los derechos de autor de Oracle. El tribunal ha considerado que la copia por parte de Google de parte del código de Java es un uso justo con base en la ponderación de los cuatro factores del *fair use*. Aunque el análisis de este caso y de los pronunciamientos del tribunal excede el ámbito de este artículo, importa señalar que, de nuevo, se ha realizado una interpretación del uso transformador flexible y favorable hacia la innovación —considerando, en particular, que la copia tenía como finalidad permitir que los programadores pudieran trabajar en un entorno de programación diferente, favoreciéndose así la interoperabilidad—.

Esta línea de precedentes se presenta, en principio, como favorable para abordar la cuestión de la copia por parte de los sistemas de IA en el marco de su proceso de aprendizaje. Siempre que la copia pueda verse como un paso intermedio para acceder a elementos no protegidos de las obras con derechos de autor (las ideas o los hechos), o como un paso intermedio para lograr un uso transformador de las obras originales, parece que la doctrina del *fair use* podrá dar cobertura a esos usos.

Sin embargo, como apunta Sobel, algunas aplicaciones del ML atentan contra las premisas sobre las que se ha basado la doctrina del «uso no expresivo» hasta ahora, lo que lleva a poner en duda si el *fair use* va a seguir amparando la innovación en esos ámbitos¹³². Como expone el autor, las dos premisas sobre las que se asentaba la doctrina del uso «no expresivo» eran: (i) que los ordenadores sólo consumían obras con una finalidad «no expresiva» y (ii) que el uso de las obras no afectaba a los mercados potenciales de una manera que se pueda considerar relevante para el *copyright*, que no protege los elementos no expresivos de las obras¹³³. El ML permite, en cambio, (i) que las máquinas utilicen las obras con un propósito «expresivo», esto es, que el fin del uso sea precisamente extraer valor de la manera en la que los autores expresan las ideas y (ii) que, como consecuencia de lo anterior, se ponga en peligro no solo el mercado potencial de la obra sino que se pueda incluso reemplazar al autor de la obra¹³⁴. Sobel llega por ello a concluir que el ML «expresivo» pone en situación de crisis a la doctrina del *fair use*, porque tanto si se considera que el uso de obras protegidas por DPI para el entrenamiento de sistemas de

¹³¹ Google LLC v. Oracle America, Inc, 553 US.

¹³² B. L. W. SOBEL, «Artificial Intelligence's...», op. cit., p. 12.

¹³³ B. L. W. SOBEL, «Artificial Intelligence's...», op. cit., p. 12.

¹³⁴ B. L. W. SOBEL, «Artificial Intelligence's...», op. cit., p. 12.

ML «expresivos» es un uso justo como si se considera que no lo es, alguno de los objetivos que se quiere conseguir con la aplicación de esa doctrina se ve afectado¹³⁵.

La solución que propone Sobel¹³⁶ es diferenciar entre los usos de las obras protegidas por DPI para el entrenamiento de IA que invaden el mercado de las obras y los que no lo invaden¹³⁷, de modo que en el caso de los usos «no invasores» se permita el aprendizaje de la IA a partir de materiales protegidos sin hacer frente al pago de ninguna licencia¹³⁸. En el caso de los usos «invasores», Sobel propone establecer un sistema equilibrado entre los distintos intereses en juego, como el de las excepciones para MTD de la DDAMUD¹³⁹.

Otros autores, como Lemley y Casey, comparten las dudas acerca de cómo evolucionará la propiedad intelectual en Estados Unidos en la materia, señalando que, a medida que las capacidades del ML aumenten, será mayor la tentación de utilizar el *copyright* como una herramienta para restringirlas¹⁴⁰ y que, en particular, existen dudas sobre si los tribunales considerarán, en todos los casos, que el aprendizaje de los sistemas de IA es un uso justo¹⁴¹. Lemley y Casey proponen incorporar en la aplicación de la doctrina del *fair use* al análisis sobre el entrenamiento de sistemas de IA un elemento adicional que denominan «*fair learning*» (aprendizaje justo), según el cual si el propósito de los sistemas de IA no es utilizar los elementos de una obra protegidos por *copyright* sino los elementos no protegidos, el uso debe en principio considerarse como justo bajo el primer factor, a no ser que el uso interfiera directamente con el mercado del titular de la obra (cuarto factor), en cuyo caso el uso no debería considerarse justo¹⁴². En el caso de que el sistema de IA utilice para el aprendizaje los elementos protegidos por *copyright*, señalan estos autores que aun así puede considerarse que el uso es justo en atención a los efectos en el mercado de la obra o en atención a su carácter transformador¹⁴³.

¹³⁵ B. L. W. SOBEL, «Artificial Intelligence's...», op. cit., p. 35.

¹³⁶ B. SOBEL, «A Taxonomy of Training Data: Disentangling the Mismatched Rights, Remedies, and Rationales for Restricting Machine Learning», *Artificial Intelligence and Intellectual Property*, (Coord. J. Lee, R. Hilty, K. Liu), Oxford University Press, Oxford, 2021, pp. 221-245.

¹³⁷ Para el autor, un uso invasor sería aquel que invade el mercado respecto del cual la propiedad intelectual concede un monopolio, en la medida en que se utiliza la expresión protegida con fines que usurpan el mercado. Un ejemplo de uso no invasor sería, como expone Sobel, el de un software que copie novelas para detectar plagios. B. SOBEL, «A Taxonomy of Training Data...», op. cit., pp. 221-245.

¹³⁸ B. SOBEL, «A Taxonomy of Training Data... », op. cit., pp. 221-245.

¹³⁹ B. SOBEL, «A Taxonomy of Training Data... », op. cit., pp. 221-245.

¹⁴⁰ M. A. LEMLEY, B. CASEY, «Fair Learning», op. cit., p. 760.

¹⁴¹ M. A. LEMLEY, B. CASEY, «Fair Learning», op. cit., p. 763.

¹⁴² M. A. LEMLEY, B. CASEY, «Fair Learning», op. cit., p. 776-777.

¹⁴³ M. A. LEMLEY, B. CASEY, «Fair Learning», op. cit., p. 777-778.

4. JAPÓN

Japón fue el primer país del mundo que modificó su ley de propiedad intelectual para incluir una excepción para las actividades de MTD, mediante una reforma operada en el año 2009¹⁴⁴. Interesa destacar que la ley de propiedad intelectual japonesa no contempla una cláusula general que abarque las excepciones y limitaciones a los DPI al estilo del *fair use* de EEUU, sino que, al igual que ocurre en el caso de la UE, lo que se regula son excepciones y limitaciones para casos concretos, que los tribunales japoneses suelen interpretar de manera flexible¹⁴⁵.

La excepción de MTD introducida en la reforma de 2009 fue considerada insuficiente para afrontar los retos que el desarrollo tecnológico planteaba, al margen de que no era clara en algunos ámbitos y creaba inseguridad jurídica¹⁴⁶. Por ello, en 2018 se procedió a una nueva reforma de ley de propiedad intelectual japonesa para introducir y modificar excepciones y limitaciones a los DPI, que entró en vigor el 1 de enero de 2019¹⁴⁷.

En lo que interesa a efectos de este estudio, nos centraremos en el análisis del artículo 30-4 de la ley japonesa, en su versión tras la modificación:

«Artículo 30-4 [Explotaciones no para disfrutar las ideas o emociones expresadas en una obra]

Está permitido explotar una obra, por cualquier medio y en la medida que se considere necesario, en cualquiera de los siguientes casos o en otros casos donde la explotación no sea para disfrutar o hacer que otra persona disfrute de las ideas o emociones expresadas en esa obra; a condición, sin embargo, de que esto no es de aplicación en el caso de que la explotación perjudique injustificadamente los intereses del titular de los derechos de autor teniendo en cuenta la naturaleza y el propósito de la obra, así como las circunstancias de la explotación:

- (i) explotaciones para utilizar la obra en experimentos para el desarrollo o puesta en práctica de tecnologías relacionadas con la grabación de sonidos e imágenes u otras explotaciones de esa obra;*

¹⁴⁴ Para más información sobre el régimen aplicable en 2009, *vid.* M. CASPERS, L. GUIBAULT, «Baseline report of policies and barriers of TDM in Europe», *Future TDM*, septiembre 2015, pp. 68-69. https://www.futuretdm.eu/wp-content/uploads/FutureTDM_D3.3-Baseline-Report-of-Policies-and-Barriers-of-TDM-in-Europe-1.pdf

¹⁴⁵ T. UENO, «The Flexible Copyright Exception for ‘Non-Enjoyment’ Purposes Recent Amendment in Japan and Its Implication», *GRUR International*, volumen 70, número 2, febrero 2021, pp. 145-148.

¹⁴⁶ M. CASPERS, L. GUIBAULT, «Baseline report...», *op. cit.*, p. 69.

¹⁴⁷ T. UENO, «The Flexible Copyright Exception...», *op. cit.*

- (ii) *explotaciones para utilizar la obra en el análisis de datos (entendiéndose por tal la extracción, comparación, clasificación u otro análisis estadístico del lenguaje, sonido, o datos de imágenes u otros elementos de los que se componga una gran cantidad de obras o un gran volumen de datos);*
- (iii) *adicionalmente a los casos expuestos en los dos apartados anteriores, explotaciones para utilizar la obra en el curso del procesamiento de datos informáticos o de otra manera que no implique percibir las expresiones de esa obra a través de los sentidos humanos (en relación con obras de programación informática, la ejecución de la obra en un ordenador debe excluirse)»¹⁴⁸.*

Como se puede comprobar, el artículo 30-4 introduce una regla general según la cual se permite la explotación de la obra cuando no se utilice con la finalidad de disfrutar de los elementos que la propiedad intelectual protege (la expresión de las ideas o emociones). La doctrina japonesa se ha referido a estos usos como de «no disfrute»¹⁴⁹.

La anterior regla viene acompañada de una importante excepción: que, aun cuando la explotación se realice con fines de «no disfrute» (de los elementos expresivos de la obra), no se permitirá su utilización si ello perjudica injustificadamente los intereses del titular de los derechos de autor, lo que habrá que evaluar en atención a la naturaleza y propósito de la obra y a las circunstancias en que se realice su explotación.

Junto a la regla general de explotación establecida en el artículo 30-4, se incluye una lista de tres ejemplos específicos de explotación de la obra con finalidad de «no disfrute»: (i) la explotación para experimentos relacionados con el desarrollo tecnológico; (ii) la explotación en el marco del análisis de datos; y (iii) la explotación de cualquier modo que no implique percibir la expresión de la obra por los sentidos humanos. La amplitud de los supuestos contemplados permite pensar no solo en la explotación de las obras en técnicas de MTD o de ML, sino en cualquier otra tecnología que se desarrolle en el futuro.

La excepción japonesa regulada en el artículo 30-4 presenta, como se puede observar, notables diferencias respecto a las excepciones para MTD contempladas en la DDAMUD.

En primer lugar, la ley japonesa no condiciona la aplicación de la excepción en función de quién sea el beneficiario de la excepción o cuál sea el uso de las obras. No hay duda, pues, de que las entidades privadas que utilicen obras

¹⁴⁸ Traducción propia a partir de la traducción al inglés. T. UENO, «The Flexible Copyright Exception...», op. cit.

¹⁴⁹ T. UENO, «The Flexible Copyright Exception...», op. cit.

protegidas por DPI para el desarrollo de sistemas de ML con fines comerciales están cubiertas por la excepción y disfrutan del mismo régimen, a estos efectos, que las entidades de investigación públicas que realicen la misma actividad con fines no comerciales. Esto supone una importante diferencia respecto de la regulación incluida en la DDAMUD, que reserva la aplicación de la excepción del artículo 3 (en la que no cabe la reserva de derechos por parte de los titulares) a las actividades de MTD realizadas por «*organismos de investigación*¹⁵⁰ e *instituciones del patrimonio cultural*¹⁵¹ con «*fines de investigación científica*».

En segundo lugar, la excepción japonesa resulta de aplicación incluso en aquellos casos en los que el titular de los DPI se haya reservado sus derechos¹⁵². Como se ha visto anteriormente, en el caso de la excepción del artículo 3 de la DDAMUD, los titulares de los derechos no pueden reservarse el uso de las obras para MTD, pero esto no ocurre así en la excepción del artículo 4 (aplicable, esencialmente, cuando la MTD se realice por entidades privadas o con fines comerciales), en la que sí se permite esa reserva. En el caso de la ley japonesa —y en coherencia con el mismo tratamiento que el artículo 30-4 otorga a las entidades privadas y públicas y a las investigaciones con fines comerciales y no comerciales— no existe esa distinción, por lo que se aplica la excepción en todo caso y con independencia de que los titulares hayan podido reservarse sus derechos.

En tercer lugar, la ley nipona no requiere que se haya accedido a la obra de manera legítima¹⁵³, lo que sí exige, como se ha visto, la DDAMUD.

Por último, el artículo 30-4 de la ley japonesa permite la explotación «*por cualquier medio*» —lo que incluiría no sólo la copia digital, sino también la analógica— y ampara no sólo la reproducción de la obra sino también la distribución y comunicación al público, lo que implica que el conjunto de datos utilizado para el desarrollo de un sistema de ML podría ser compartido¹⁵⁴.

Importa señalar que, aunque la redacción del artículo 30-4 de la ley japonesa y la doctrina del uso con finalidad de «no disfrute» que en él se consagra podrían recordar a la doctrina del uso «no expresivo» desarrollada en los tribunales de EEUU a partir de *Sega v. Accolade*, no es posible establecer un paralelismo que prescinda, sin más, de otras consideraciones. Como se ha visto anteriormente, con cita de Sobel, la doctrina del uso «no expresivo» se ha aplicado principal-

¹⁵⁰ La definición de «*organismo de investigación*» a efectos de la DDAMUD se incluye en el apartado 1 del artículo 2. *Vid* nota 94 anterior.

¹⁵¹ La definición de «*institución responsable del patrimonio cultural*» a efectos de la DDAMUD se incluye en el apartado 3 del artículo 2. *Vid* nota 95 anterior.

¹⁵² T. UENO, «The Flexible Copyright Exception...», op. cit.

¹⁵³ T. UENO, «The Flexible Copyright Exception...», op. cit.

¹⁵⁴ T. UENO, «The Flexible Copyright Exception...», op. cit.

mente en casos donde no se extraía valor de los elementos expresivos de las obras y donde no se planteaba el riesgo de afectación del mercado potencial de un modo tal que no sólo la obra copiada, sino el propio autor de dicha obra, pudieran verse reemplazados. En este sentido, no hay que olvidar que la doctrina del uso «no expresivo» se ha desarrollado en el marco de la aplicación del *fair use*, que exige la ponderación de los cuatro factores estudiados más arriba, entre los que se incluyen el propósito y el carácter del uso (primer factor) y el efecto del uso sobre el mercado potencial o sobre el valor de la obra protegida (cuarto factor). Y si bien es posible que los tribunales estadounidenses apliquen el *fair use* de un modo tal que dé lugar a resultados similares a los que prevé el artículo 30-4 de la ley nipona, también es posible que, como se ha visto más arriba, adopten soluciones distintas en aquellos casos donde el ML vaya dirigido a extraer valor de los elementos expresivos de la obra con posible afectación de su mercado potencial.

En el caso del artículo 30-4 de la ley nipona, el legislador ya ha realizado esa ponderación en los tres supuestos específicos de uso con finalidad de «no disfrute» que se regulan en el mismo, optando por permitir la explotación de las obras en esos supuestos. Como señala UENO, al amparo de la excepción contemplada en el apartado (ii) del artículo 30-4, no habría problema para copiar libros, periódicos o *tweets* con el propósito de predecir tendencias de mercado con fines comerciales, ni tampoco para copiar las películas de Star Wars o las canciones de los Beatles a fin de desarrollar un modelo de ML que genere una nueva obra replicando el estilo de las anteriores¹⁵⁵. Como explica el autor, la regulación del artículo 30-4 se justifica en que sólo la explotación de una obra con finalidad de disfrute debe estar cubierta por la propiedad intelectual, mientras que si se explota con un propósito de «no disfrute», no existe razón para que los titulares de los DPI reciban compensación y la propiedad intelectual no debe cubrir esa explotación¹⁵⁶.

La reforma de la ley japonesa supone, pues, un impulso muy decidido a la innovación, siendo en la actualidad una de las jurisdicciones donde la utilización de materiales protegidos por DPI para el desarrollo de sistemas de IA se presenta como más favorable, no sólo por la amplitud con la que se ha formulado la excepción sino, también, por la seguridad jurídica existente en torno a estas actividades —en contraste, por ejemplo, con el caso de EEUU, que no cuenta con una norma ni con un precedente claro que aporte a las empresas que desarrollan IA el mismo grado de certidumbre que en Japón—. Por ello, se habla de Japón como un «paraíso» para la MDT y el ML¹⁵⁷.

¹⁵⁵ T. UENO, «The Flexible Copyright Exception...», op. cit.

¹⁵⁶ T. UENO, «The Flexible Copyright Exception...», op. cit.

¹⁵⁷ T. UENO, «The Flexible Copyright Exception...», op. cit.

V. REFLEXIONES A LA LUZ DE LAS DISTINTAS SOLUCIONES ADOPTADAS EN LA UE, EEUU Y JAPÓN. UNA MIRADA HACIA EL FUTURO DE LA PROPIEDAD INTELECTUAL ANTE LOS RETOS DEL ML

En el apartado anterior se han expuesto tres soluciones completamente distintas para abordar la cuestión que es objeto de estudio: (i) la solución europea, que ha introducido excepciones a la protección de los DPI para actividades de MTD, pero que sólo es imperativa para los particulares cuando la beneficiaria es una entidad de investigación de las contempladas en la DDAMUD¹⁵⁸ y la investigación no se realiza con fines comerciales; (ii) la solución estadounidense, que mediante *case law* ha ido configurando una doctrina del *fair use* que hasta ahora se ha mostrado favorable a las copias realizadas con propósito «no expresivo» o utilizadas de manera transformadora; y (iii) la solución japonesa, que resuelve la tensión entre la innovación y la propiedad intelectual introduciendo una excepción a la protección de los DPI para el entrenamiento de los sistemas de IA que es aplicable a todo tipo de beneficiarios y a usos comerciales y no comerciales, sin posibilidad en ningún caso de reserva de derechos por parte de los titulares de los DPI.

En nuestra opinión, este análisis comparado es particularmente útil para delimitar los contornos de la discusión jurídica e identificar algunas claves que deben ser tenidas en cuenta en los desarrollos legislativos que seguirán produciéndose en los próximos años. Ante una cuestión tan compleja como la planteada, que implica la conciliación de importantes intereses en juego, parece relevante ir asentando bases sólidas sobre las que avanzar en el diálogo, y sin duda la experiencia comparada ofrece ya valiosas enseñanzas. Ha de recordarse que, en pleno debate sobre la cuestión —como se ha expuesto al inicio, la OMPI ha lanzado una consulta pública sobre IA recientemente en la que se incluye esta materia¹⁵⁹—, difícilmente las soluciones alcanzadas hasta la fecha pueden considerarse definitivas. Al menos en aquellas jurisdicciones como la UE y EEUU, donde el ámbito de la discusión jurídica (en forma legislativa, en la primera, y en forma jurisprudencial, en la segunda) no ha abarcado en su totalidad los retos que supone el ML (particularmente, cuando se copian elementos «expresivos» de las obras) para la propiedad intelectual.

En las páginas que siguen expondremos algunas de esas claves que, a juicio de quien escribe, deberían ser tenidas en cuenta para construir el marco jurídico del desarrollo de la IA desde la perspectiva de la propiedad intelectual.

¹⁵⁸ *Vid.* definición de «organismo de investigación» y de «institución responsable del patrimonio cultural» incluidas en los apartados 1 y 3 del artículo 2 de la DDAMUD, *vid.* notas 94 y 95.

¹⁵⁹ WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI), Second Session, *Draft Issues Paper on Intellectual Property and Artificial Intelligence*, 13 diciembre 2019, p. 5.

Avanzamos ya que no nos pronunciaremos de un modo definitivo sobre cuál sería la solución ideal o perfecta al dilema que se plantea en el título de este artículo, y ello por dos razones.

Primero, porque es una cuestión compleja cuya correcta valoración requiere estudiar en mayor profundidad y tomar en consideración el proceso completo de creación de valor a través de ML y su diálogo con la propiedad intelectual. Esto es, no solo deben considerarse los *inputs* que se incorporan a ese proceso (que es lo que aquí se estudia), sino también los *outputs* obtenidos (las «obras» creadas por IA) y el propio proceso de creación (la posible protección de los propios modelos de ML)¹⁶⁰. En la configuración del régimen legal aplicable debe ser tenida en cuenta también la posible interacción con la normativa de protección de secretos empresariales¹⁶¹.

Y segundo, porque la realidad es que no existe en esta materia la solución perfecta. La tensión que se plantea entre la innovación y la propiedad intelectual cuando de entrenar a la IA se trata sólo puede resolverse definiendo un equilibrio entre los intereses en juego que no tiene por qué coincidir —y de hecho no coincide en la actualidad— en las distintas jurisdicciones. El balance que se adopte en cada país será, al final, un reflejo de sus sistemas políticos, económicos y sociales y del rol que quiera jugar en esta partida de ajedrez global que es la carrera hacia el liderazgo en IA.

Hechas estas consideraciones, pasamos a exponer algunas de las claves que estimamos que deberían ser tenidas en cuenta para dar respuesta a la cuestión objeto de estudio desde una perspectiva de *lege ferenda*. El enfoque de estas consideraciones no está centrado en ninguna jurisdicción en particular —dado el ámbito global de la discusión iniciado en el seno de la OMPI—, si bien se harán referencias concretas a las regulaciones específicas cuando sea preciso.

¹⁶⁰ Este es el enfoque que sigue el Instituto Max Planck en su reciente publicación sobre la cuestión estudiada. J. DREXL, R. M. HILTY, L. DESAUNETTES-BARBERO, J. GLOBOCNIK, B. GONZÁLEZ OTERO, J. HOFFMAN, D. KIM, S. KULHARI, H. RICHTER, S. SCHEUERER, P. R. SLOWINSKI, K. WIEDEMANN, «Artificial Intelligence and Intellectual Property Law Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate», *Max Planck Institute for Innovation and Competition Research Paper No 21-10*, abril 2021, pp. 1-2.

¹⁶¹ Como explican J. DREXL y otros, en el caso de la UE podría darse el caso de que los datos obtenidos mediante MTD podrían ser considerados secretos comerciales en el sentido de la Directiva 2016/943 del Parlamento Europeo y del Consejo de 8 de junio de 2016 relativa a la protección de los conocimientos técnicos y la información empresarial no divulgados (secretos comerciales) contra su obtención, utilización y revelación ilícitas, si el acceso a los documentos sobre los que se ha aplicado MTD está restringido contractualmente o mediante medidas técnicas de protección. J. DREXL *et ál.*, «Artificial Intelligence and Intellectual Property Law...», op. cit, p. 8.

1. COMPLEJIDAD DE LA CUESTIÓN E INSUFICIENCIA DE LA DICOTOMÍA IDEA-EXPRESIÓN PARA DAR RESPUESTA A LOS RETOS QUE PLANTEA EL ENTRENAMIENTO DEL ML PARA LA PROPIEDAD INTELECTUAL

El ML, como se ha expuesto, introduce una complejidad adicional que no siempre se ha tenido en cuenta en el debate en torno al uso de materiales protegidos por la propiedad intelectual para MTD, en el que la doctrina abogaba mayoritariamente por tolerar ese uso partiendo de la dicotomía idea-expresión o hecho-expresión.

El argumento habitualmente utilizado de que debe permitirse que los sistemas de IA copien porque es su manera de acceder a contenido no protegido por la propiedad intelectual (las ideas y los hechos), de la misma manera que se permite que las personas lean textos para extraer las ideas y los hechos, obvia que, en muchas ocasiones, los sistemas de ML sí que están «interesados» en copiar lo que la propiedad intelectual protege: los elementos expresivos de la obra o, dicho de otro modo, la manera en la que se expresan las ideas, hechos, emociones, opiniones o pensamientos de los autores. De hecho, el desarrollo de modelos de ML que son capaces de realizar actividades artísticas (componer canciones, pintar cuadros, etc.) a partir del estudio de los elementos expresivos de obras y prestaciones utilizadas en su entrenamiento ha evolucionado considerablemente en los últimos años y es previsible que continúe en la misma línea.

En atención a lo anterior, ha de concluirse que la dicotomía idea-expresión no sirve para resolver por sí misma el reto que plantea el ML para la propiedad intelectual, o al menos no lo hace siguiendo la línea argumental tradicionalmente esgrimida al aplicar este principio al aprendizaje de la IA.

Una posibilidad para seguir profundizando en esa línea argumental sería plantearse, al hilo de la dicotomía idea-expresión, si la propiedad intelectual protege en todo caso la expresión de las ideas, o la protege solo en tanto en cuanto el destinatario de esa expresión, quien la «disfrute» —utilizando la terminología de la ley nipona— sea un ser humano. Si se parte de la segunda tesis, como hace la ley japonesa, podría concluirse que no es preciso tratar de forma diferenciada los casos en que los sistemas de ML «aprenden» de la propia expresión de las obras de aquellos casos en que «aprenden» de las ideas que extraen de las obras. Y, por ende, concluir con base en la dicotomía idea-expresión, que debe permitirse que los sistemas de IA sean entrenados con obras protegidas por DPI en cualquier supuesto, con independencia de si copian la expresión para acceder a las ideas o copian la expresión para extraer valor de ella, pues en ningún caso quien «disfruta» del uso de las obras es un ser humano.

Esta formulación de la dicotomía idea-expresión para resolver la cuestión, sin embargo, no aborda el desafío principal que presenta el ML que es entrenado a partir de los elementos expresivos, que es la posible afectación al mercado de la

obra copiada o a su valor, e incluso la aparición de nuevos «creadores» artificiales que hagan la competencia directa a los autores de los que han «aprendido». Nótese, en esta línea, que incluso la ley japonesa contiene, en la cláusula general del artículo 30-4, una salvaguarda específica para la aplicación de la excepción general que en ella se contiene: que, aun cuando la explotación se realice con fines de «no disfrute» (de los elementos expresivos de la obra), no se permitirá la utilización de la obra si ello perjudica injustificadamente los intereses del titular de los derechos de autor, lo que habrá que evaluar en atención a la naturaleza y propósito de la obra y a las circunstancias en que se realice su explotación.

Por todo lo anterior, al abordar el estudio de los retos que plantea el entrenamiento de sistemas de IA mediante materiales protegidos por DPI, hay que partir de que el ML ha introducido una dificultad añadida en el debate que hace que no sea posible resolver el problema aludiendo, con base en la dicotomía idea-expresión, a que la copia sólo tiene por objeto acceder a las ideas de las obras, porque esa premisa sencillamente no se cumple en todos los casos. Habrá, pues, que integrar más elementos en el análisis.

Lo anterior no significa que la dicotomía idea-expresión no sea útil para resolver la cuestión que nos ocupa: lo será para abordar aquellos casos en los que la copia tiene como finalidad acceder a la parte de la obra que la propiedad intelectual no protege (las ideas, los hechos o la información). Para los casos en que la copia tenga como finalidad acceder a lo que la propiedad intelectual sí que protege (la expresión de esas ideas, hechos o información), será necesario efectuar consideraciones adicionales ante la complejidad añadida que se plantea.

Queremos aclarar que no se está aquí afirmando que, necesariamente, haya de darse en el plano legislativo una solución distinta a los casos de aprendizaje de IA a partir de la copia de obras protegidas para acceder a las ideas o los hechos y a los casos en los que se copia para acceder a la expresión —en los siguientes apartados se estudiará cada una de estas cuestiones de manera específica y se verán alternativas—. Pero a efectos de la discusión jurídica sobre la materia que siente las bases para los futuros desarrollos legislativos, es importante contemplar y analizar de forma separada estos dos tipos de usos de las obras utilizadas en el entrenamiento de sistemas de ML en cuanto plantean problemas diferentes.

2. CASOS DONDE LAS OBRAS Y PRESTACIONES PROTEGIDAS POR DPI SE UTILIZAN EN EL APRENDIZAJE DE LA IA COMO MEDIO PARA ACCEDER A LAS IDEAS O LOS HECHOS

En los supuestos donde los sistemas de IA son entrenados con obras y prestaciones protegidos por DPI como medio de acceder a la parte no protegida por

la propiedad intelectual (las ideas, los hechos, la información), es coherente considerar que la utilización de esos materiales sin autorización de los titulares de los derechos exclusivos no debe ser considerada una infracción con base en la dicotomía idea-expresión.

En este ámbito entrarían todos aquellos supuestos de ML —que son una parte muy significativa, por no decir la mayoría— donde se entrena al algoritmo para extraer ideas o conceptos a partir de los datos sin que se extraiga un valor de la expresión, como ocurre con los sistemas de reconocimiento de imagen o los traductores de textos.

El estudio comparado que se ha efectuado en la sección anterior de este trabajo permite afirmar que existe cierto consenso en torno a esta solución incluso en jurisdicciones con tradiciones muy distintas, pues definitivamente la ley japonesa la adopta, el *case law* en EEUU se ha desarrollado en esta línea con la doctrina del uso «no expresivo» y en el ámbito de la Unión Europea la doctrina se ha mostrado también decididamente favorable a esta tesis desde que, hace ya tiempo, defendiera en el ámbito de la MDT que «el derecho a leer debería ser el derecho a minar».

Quien no ha compartido esta visión ha sido, hasta ahora, el legislador europeo al configurar el limitado alcance de las excepciones de MDT sin establecer una regulación más flexible para los casos donde el uso de las obras y prestaciones tenga como finalidad acceder a la información incluida en esos materiales y la copia sea un mero requisito tecnológico para ello. En este contexto, parece claro que si la UE no emprende reformas legislativas para ampliar el alcance de las excepciones de la DDAMUD de modo que se permita el uso de obras y prestaciones protegidas por DPI para MDT y para ML con finalidad de acceder a los elementos no protegidos, con independencia de quién sea la entidad que realice esas actividades (entidades públicas o privadas) y de que el fin sea comercial o no comercial, se producirá un freno importante a la innovación en los Estados miembros.

En conclusión, no se encuentran en la perspectiva comparada argumentos para impedir la utilización de materiales protegidos por DPI para el entrenamiento de IA cuando lo que se persigue es extraer ideas, hechos, información o conceptos de esos materiales. Aunque existen razones para sostener que ese uso quedaría, incluso, fuera del alcance de la propiedad intelectual (*vid.* sección III anterior de este trabajo), consideramos más adecuado, a fin de proporcionar seguridad jurídica sobre la cuestión y de evitar que medidas tecnológicas o contractuales de protección puedan constituir un impedimento, que se regule la cuestión por vía de excepciones y limitaciones a los DPI.

3. CASOS DONDE LAS OBRAS Y PRESTACIONES PROTEGIDAS POR DPI SE UTILIZAN EN EL APRENDIZAJE DE LA IA CON EL PROPÓSITO DE EXTRAER VALOR DE LOS ELEMENTOS EXPRESIVOS

En los casos donde los sistemas de IA son entrenados con obras y prestaciones protegidos por DPI con el propósito de aprovechar el valor de la parte protegida por la propiedad intelectual (la expresión de las ideas o los hechos), para decidir si la copia debe ser considerada una infracción será necesario un mayor análisis, en tanto que la cuestión no puede solventarse fácilmente atendiendo únicamente al principio de la dicotomía idea-expresión, como se ha explicado en el apartado 1 anterior.

El desafío principal que presenta el ML que copia elementos expresivos es la posible afectación al mercado de la obra copiada o a su valor, e incluso la aparición de nuevos «creadores» artificiales que hagan la competencia directa a los autores humanos de los que han «aprendido». Para ilustrarlo con un ejemplo, basta imaginar que, inspirada por el éxito de «El Nuevo Rembrandt», una compañía decide lanzar un proyecto similar pero basado en obras de un reconocido pintor que esté actualmente en activo, siendo capaz de crear un sistema de IA que «pinte» obras del mismo estilo que ese pintor. O, por pensar en otro supuesto, imagínese que se entrena a un sistema de IA con novelas de un conocido autor de modo que sea capaz de producir nuevas novelas imitando su estilo. Por plantear un último caso, imagínese que se entrena a un sistema de IA con las cien canciones más populares del momento de un determinado estilo musical, de distintos artistas, de modo que sea capaz de crear canciones «exitosas» en ese estilo pero sin imitar a ningún cantante en particular. ¿Debería permitirse en todos estos casos la utilización de las pinturas, de las novelas y de las canciones para entrenar a los sistemas de IA sin el consentimiento de los titulares de los DPI y sin el pago de ninguna compensación económica?

La cuestión es ciertamente compleja y requiere de un estudio en mayor profundidad del que se puede realizar en estas páginas, en las que sólo se expondrán algunas de las alternativas planteadas.

Una primera posibilidad sería entender que también en estos casos en los que la copia tiene como finalidad los elementos expresivos ha de permitirse que los sistemas de IA «aprendan» con estas obras y prestaciones protegidas por DPI sin autorización de los titulares de los derechos y sin pago de compensación. A este camino aboca la tesis japonesa con la tesis del «no disfrute», ya expuesta.

Otro argumento para sustentar esta posición sería asimilar el aprendizaje de la IA al proceso de inspiración y creación humana. Podría alegarse que, del mismo modo que se permite que los autores se inspiren a partir de las obras de otros autores y sin duda se trata de un uso permitido por la propiedad intelectual (por ejemplo, un escritor que se forma e inspira a partir de las novelas

que lee), debería permitirse que los sistemas de IA hagan lo propio. Esta es la tesis que defiende Chiou, para quien no debería haber diferencias entre el proceso de inspiración humana y el proceso de «inspiración» de la IA y, por ello, concluye que esa fase de «inspiración» de la IA debe quedar fuera del alcance de la propiedad intelectual¹⁶². Sostiene este autor que si lo que produce el sistema de IA tras completar la fase de aprendizaje puede considerarse una copia o transformación de las obras utilizadas para el entrenamiento del algoritmo, será entonces un problema del *output* y no del *input*, y lo que deberá controlarse desde el prisma de la propiedad intelectual es ese *output*¹⁶³ sin que ello condicione la licitud de utilizar esas obras en la fase de entrenamiento.

Se pueden realizar, sin embargo, algunas objeciones a la tesis de asimilar la inspiración y creación humana y la que se atribuye a la IA como vía para permitir el uso de materiales protegidos por DPI para el entrenamiento de la IA. La primera es que, como apunta Sobel, que un autor humano se inspire con ideas extraídas de obras es un uso permitido por la propiedad intelectual, pero eso no le exime de pagar por las obras que consume, por más que luego aproveche ese conocimiento para crear nuevas obras¹⁶⁴. Habría que cuestionarse, por tanto, las razones para otorgar un trato más favorable a la IA que a los autores humanos.

Otra objeción al planteamiento de asimilar la «inspiración» de la IA a la humana es que el *output* que genera un autor humano y el que genera la IA no son, al menos por el momento, asimilables desde el punto de vista de la propiedad intelectual. Un escritor humano no puede evitar que otros autores se inspiren con sus novelas, pero sí dispone de herramientas en el ámbito de la propiedad intelectual para dirigirse frente a las obras que creen esos mismos autores si, en un «exceso» de inspiración, generan una novela que sea demasiado similar a la suya. Esta es la manera en la que la propiedad intelectual salvaguarda adecuadamente los intereses de la sociedad y los del autor en cuestión: la «inspiración» queda fuera del ámbito de la propiedad intelectual, lo que se crea después no. Sin embargo, en pleno debate sobre si la propiedad intelectual debe de algún modo proteger las creaciones generadas mediante IA o si estas deben integrarse en el dominio público¹⁶⁵, este modelo de inspiración-creación

¹⁶² T. CHIOU, «Copyright Law and Algorithmic Creativity...», op. cit., pp. 286-287.

¹⁶³ T. CHIOU, «Copyright Law and Algorithmic Creativity...», op. cit., pp. 286-287.

¹⁶⁴ Así lo señala Sobel, al comparar el caso de un profesor que estudia obras protegidas por DPI que le inspiran para crear una nueva obra y el del ML, y concluir que si la productividad futura no excusa el uso no autorizado de obras por humanos, tampoco debería excusar ese consumo de obras por parte de la IA. B. L. W. SOBEL, «Artificial Intelligence's...», op. cit., p. 26. Ha de tenerse en cuenta, sin embargo, que existe también la posibilidad de establecer que las máquinas «paguen» exactamente lo mismo que un autor humano a efectos de inspirarse, en lugar de tener que pagar una licencia para ML.

¹⁶⁵ Así lo entiende, entre otros, Fernández Carballo-Calero. P. FERNÁNDEZ CARBALLO-CALERO, *La propiedad intelectual de las obras creadas por inteligencia artificial*, Aranzadi, Cizur-Menor (Navarra), 2021.

humana con control únicamente en la fase de *output* no es, al menos por ahora, replicable. No existe en este momento certeza sobre el tratamiento que se le va a dar a esa novela producida por la IA que podría considerarse una copia o una transformación de la utilizada en la fase de entrenamiento, por lo que si sólo se buscan soluciones a la cuestión planteada desde la perspectiva del *output* producido, la solución alcanzada puede resultar inadecuada o insuficiente.

Se podría también objetar, frente al planteamiento de equiparar la «inspiración» de la IA a la humana, algo que, en el marco del proceso paulatino e inconsciente de atribuir facultades humanas a las máquinas —a lo que también hemos contribuido en este artículo, al decir que las máquinas «aprenden»—, a veces se pasa por alto: la «inspiración» de la IA no tiene nada que ver con la inspiración humana, ni tampoco el proceso creativo. Sin ánimo de adentrarse ahora en consideraciones filosóficas, la realidad es que lo que las personas crean no es sólo el resultado de las obras que han «consumido» como inspiración, sino que es también el resultado de su personalidad, sus vivencias, opiniones y creencias. La «inspiración» de la IA y el proceso de creación se parece mucho más a un proceso productivo: a partir de unos determinados *inputs* (las obras utilizadas para entrenar al algoritmo), se generan unos *outputs* (las nuevas obras que produce el sistema), que además pueden producirse a una velocidad y volumen fuera del alcance de cualquier autor humano. Esta correlación *input-output* y el potencial de producción a escala industrial del *output* constituyen importantes diferencias entre el proceso de creación humana y el de la IA sobre las que es preciso reflexionar al diseñar las posibles soluciones.

Otra línea de razonamiento distinta para dar respuesta a la cuestión de si debe permitirse el entrenamiento de sistemas de IA con obras protegidas por DPI con el propósito de que estos sistemas aprendan de la expresión es condicionar la licitud del uso a cuál sea el *output* que producen; en otras palabras, a qué es lo que hace la IA a partir de los datos de entrenamiento. Esta es la posición de Sobel que, como se ha expuesto antes, propone diferenciar entre los usos de las obras protegidas por DPI para el entrenamiento de IA que invaden el mercado de las obras de los que no lo invaden (un uso «invasor» sería aquel que invade el mercado respecto del cual la propiedad intelectual concede un monopolio, en la medida en que se utiliza la expresión protegida con fines que usurpan el mercado)¹⁶⁶. Sobel sostiene que ha de permitirse que los sistemas de IA aprendan a partir de materiales protegidos por la propiedad intelectual cuando el resultado del entrenamiento no dé lugar a un uso invasor del mercado (como sería el de un software que copie novelas para detectar plagios)¹⁶⁷. En el caso de los usos «invasores», Sobel sugiere establecer un sistema equilibrado

¹⁶⁶ B. SOBEL, «A Taxonomy of Training Data: Disentangling the Mismatched Rights, Remedies, and Rationales for Restricting Machine Learning», *Artificial Intelligence and Intellectual Property*, (Coord. J. Lee, R. Hilty, K. Liu), Oxford University Press, Oxford, 2021, pp. 221-245.

¹⁶⁷ B. SOBEL, «A Taxonomy of Training Data... », op. cit., pp. 221-245.

entre los distintos intereses en juego, para el que las excepciones para MTD de la DDAMUD podrían servir de modelo, con el añadido de que habría que regular específicamente aquellos casos en los que los titulares de DPI sí que están dispuestos a permitir que se utilicen las obras y prestaciones para el desarrollo de sistemas de IA a cambio de una remuneración¹⁶⁸.

Otra posibilidad para regular la cuestión planteada consistiría, en relación con la expuesta en el párrafo anterior, en permitir en todos los casos la utilización de materiales protegidos por DPI para el aprendizaje de la IA incluso cuando se persiga extraer valor de los elementos expresivos, pero estableciendo al mismo tiempo un derecho de remuneración en algunos casos en que pueda considerarse adecuado, por ejemplo cuando lo que produzca la IA tenga fines comerciales¹⁶⁹. En este escenario, para evitar los altos costes de transacción en la negociación de licencias, sería recomendable intentar fomentar modelos de ventanilla única (*one-stop shop*) donde las entidades que desarrollen IA puedan obtener las licencias correspondientes de manera eficiente¹⁷⁰.

Una alternativa adicional sería que los casos donde se entrena a la IA con materiales protegidos por DPI para extraer valor de la expresión se aborden mediante la aplicación de una cláusula flexible, al estilo del *fair use*, que considere que el uso es justo o no atendiendo a ciertos elementos, como el carácter comercial del uso o el efecto sobre el valor de mercado de las obras. Esta es la solución que proponen Lemley y Casey, como se ha visto antes, al incorporar a la doctrina del *fair use* el elemento del «*fair learning*» (aprendizaje justo), según el cual si el propósito de los sistemas de IA no es utilizar los elementos de una obra protegidos por *copyright* sino los elementos no protegidos, el uso debe en principio considerarse como justo, a no ser que interfiera directamente con el mercado del titular de la obra¹⁷¹. En el caso de que el sistema de IA utilice para el aprendizaje los elementos protegidos por *copyright*, aun así podría considerarse que el uso es justo en atención a los efectos en el mercado de la obra o en atención a su carácter transformador¹⁷².

En conclusión, el uso de materiales protegidos por DPI para el aprendizaje de los elementos expresivos por parte de la IA plantea, por el momento, muchos interrogantes sobre cómo proteger adecuadamente los intereses de los titulares de los derechos exclusivos ante la amenaza que puede suponer para el mercado de sus obras o para su valor la nueva «producción» que genere la IA. Es una

¹⁶⁸ B. SOBEL, «A Taxonomy of Training Data... », op. cit., pp. 221-245.

¹⁶⁹ C. GEIGER, «The Missing Goal-Scorers...», op. cit., p. 9 nota 33,

¹⁷⁰ Sobre esta cuestión, *vid.* M. SENFTLEBEN, T. MARGONI, D. ANTAL, B. BODO, S. VAN GOMPEL, C. HANDKE, M. KRETSCHMER, J. POORT, J. QUINTAIS, S. SCHWEMER, *Ensuring the Visibility and Accessibility of European Creative Content on the World Market: The Need for Copyright Data Improvement in the Light of New Technologies*, 2021.

¹⁷¹ M. A. LEMLEY, B. CASEY, «Fair Learning», op. cit., pp. 776-777.

¹⁷² M. A. LEMLEY, B. CASEY, «Fair Learning», op. cit., pp. 777-778.

cuestión sobre la que no existe consenso en la doctrina y que no sería descartable que sea definitivamente resuelta de modo distinto en cada jurisdicción, en función de cuáles sean los intereses en conflicto cuya protección se prefiera primar.

En todo caso, para construir el marco jurídico aplicable a este tipo de entrenamiento de la IA será relevante atender a cuál sea el tratamiento que se otorgue desde la propiedad intelectual a las «creaciones» que produzcan los sistemas de IA a partir de los datos de entrenamiento, por las razones antes expuestas al tratar sobre el control del *input* y del *output* del proceso productivo de la IA.

4. OTRAS CONSIDERACIONES

En este apartado se apuntarán algunas consideraciones adicionales que han de tenerse en cuenta al regular la cuestión que es objeto de estudio y que, ante la imposibilidad de abordar aquí, dejamos enunciadas:

- (i) Debe revisarse si tiene sentido exigir que se haya obtenido acceso lícito a las obras o prestaciones en las que se incluyen los datos para que resulten aplicables las excepciones y limitaciones a los DPI —como establece, por ejemplo, la DDAMUD al regular las excepciones para MTD—, a la luz de la protección de los intereses de los titulares y del impacto en la innovación¹⁷³.
- (ii) Se ha planteado la cuestión de si el aprendizaje de un modelo de ML podría afectar al derecho de los autores a impedir cualquier deformación, mutilación u otra modificación de la obra que cause perjuicio a su honor o reputación, en los términos del artículo 6 bis de la Convención de Berna¹⁷⁴. Se trata de una cuestión que requeriría un estudio particular. En todo caso, para evitar cualquier duda al respecto, sería conveniente que, al regular en el plano legislativo las excepciones y limitaciones para el entrenamiento de modelos de ML, se aclarara la cuestión para evitar un escenario de inseguridad jurídica que frene la innovación.
- (iii) Es preciso profundizar en el estudio de la posibilidad de protección de bases de datos creadas específicamente para el entrenamiento de sistemas de IA desde el prisma del derecho *sui generis*¹⁷⁵, en atención a las

¹⁷³ S. FLYN, C. GEIGER, J. QUINTAIS, T. MARGONI, M. SAG, L. GUIBAULT, M. CARROLL, «Implementing User Rights for Research in the Field of Artificial Intelligence», *Washington College of Law Research Paper No. 2020-12*, pp. 7-8.

¹⁷⁴ J. DREXL *et ál.*, «Artificial Intelligence and Intellectual Property Law Position Statement of the Max Planck Institute...», op. cit., p. 12.

¹⁷⁵ Artículo 7.1 de la 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.

- posibles inversiones efectuadas en adquisición de los datos a terceros o en el etiquetado de los datos¹⁷⁶.
- (iv) Para que las excepciones y limitaciones en materia de aprendizaje de sistemas de IA favorezcan la innovación, es preciso que se regulen de manera efectiva las barreras contractuales y tecnológicas que puedan establecer los titulares de los DPI para el acceso a los materiales protegidos por la propiedad intelectual. O, dicho en otras palabras: que si el legislador permite que en determinados casos que las obras y prestaciones sean utilizadas para entrenar IA sin ninguna posibilidad de reserva de derechos por parte de los titulares, no puedan éstos impedirlo mediante disposiciones contractuales o medidas tecnológicas de protección.
- (v) A fin de evitar que la utilización de materiales no protegidos por DPI pueda acabar estando más restringida que la de materiales protegidos por DPI por la aplicación de medidas contractuales o tecnológicas (*vid.*, a este respecto, lo expuesto al tratar la sentencia del TJUE del caso *Ryanair*¹⁷⁷, en la sección IV de este artículo), parece recomendable que la legislación que introduzca, en su caso, excepciones y limitaciones a los DPI para el aprendizaje de la IA, se encargue también de abordar específicamente el uso de materiales no protegidos por la propiedad intelectual para el mismo fin.
- (vi) Como argumento en favor de permitir el entrenamiento de sistemas de IA con materiales protegidos por la propiedad intelectual sin la autorización de sus titulares, se ha esgrimido que de ese modo se incentiva a los desarrolladores de IA a emplear conjuntos de datos de entrenamiento de mayor cantidad, variedad y calidad, lo que redundará en una IA más justa que evite resultados sesgados o tenga consecuencias discriminatorias. Sin negar el razonamiento, es preciso también reflexionar sobre si la consecución de ese objetivo de lograr una IA respetuosa con los derechos fundamentales ha de venir impulsada desde la propiedad intelectual, siendo posible la promulgación de normas que vayan específicamente dirigidas a alcanzar ese objetivo en el marco de una regulación más amplia de la IA¹⁷⁸.
- (vii) En la configuración de las excepciones y limitaciones en materia de aprendizaje de sistemas de IA ha de tenerse siempre presente el impacto de la regulación en la innovación y en quiénes habrán de ser sus impulsores (entidades públicas o privadas, y dentro de éstas, todos los segmentos empresariales o sólo las grandes corporaciones). En aquellas jurisdicciones en las que se establezca que el uso de obras y

¹⁷⁶ J. DREXL *et ál.*, «Artificial Intelligence and Intellectual Property Law Position Statement of the Max Planck Institute...», op. cit., pp. 6-7.

¹⁷⁷ B. GONZÁLEZ OTERO, «Las excepciones de minería de textos y datos...», op. cit, p. 20.

¹⁷⁸ *Vid.*, por ejemplo, la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión.

prestaciones protegidas por la propiedad intelectual para el desarrollo de IA requiere del pago de licencia, la capacidad de innovar quedará principalmente a cargo de las grandes corporaciones y se hará muy difícil que las pequeñas y medianas empresas —segmento en el que se incluyen las *start-up* tecnológicas— puedan competir, lo que podrá provocar una fuga de talento de esas empresas hacia otras jurisdicciones.

VI. CONCLUSIONES

La utilización de obras y prestaciones protegidas por DPI para el desarrollo de sistemas de IA (en particular, de ML) sin el consentimiento de los titulares de los derechos plantea un desafío para la propiedad intelectual. Ello es así en la medida en que, durante el proceso de entrenamiento de los algoritmos, tienen lugar varias actividades que afectan a los derechos exclusivos, principalmente el derecho de reproducción, si bien, de entender que esas actividades constituyen una infracción de los DPI, la propiedad intelectual se situaría como un obstáculo a la innovación y al progreso científico.

La tensión entre la protección de los DPI y el impulso a la innovación ha sido resuelta de manera muy distinta en las jurisdicciones en que se ha abordado la cuestión. Sirven de ejemplo los casos de la UE, EEUU y Japón, que ofrecen tres modelos distintos de solución: (i) la solución europea, que ha introducido excepciones a la protección de los DPI para actividades de MTD, pero que sólo es imperativa para los particulares cuando la beneficiaria es una entidad de investigación de las contempladas en la DDAMUD¹⁷⁹ y la investigación no se realiza con fines comerciales; (ii) la solución estadounidense, que mediante *case law* ha ido configurando una doctrina del *fair use* que hasta ahora se ha mostrado favorable a las copias realizadas con propósito «no expresivo» o utilizadas de manera transformadora; y (iii) la solución japonesa, que resuelve la tensión entre la innovación y la propiedad intelectual introduciendo una excepción a la protección de los DPI para el entrenamiento de los sistemas de IA que es aplicable a todo tipo de beneficiarios y a usos comerciales y no comerciales, sin posibilidad en ningún caso de reserva de derechos por parte de los titulares de los DPI.

La comparación de estos tres modelos es útil para el estudio de nuevos desarrollos legislativos sobre la cuestión, que a buen seguro se producirán en un futuro próximo, al menos en la UE, ante los frenos que la regulación actual impone para el progreso tecnológico.

¹⁷⁹ Vid. definición de «organismo de investigación» y de «institución responsable del patrimonio cultural» incluidas en los apartados 1 y 3 del artículo 2 de la DDAMUD.

Desde una perspectiva de *lege ferenda*, se considera que la discusión jurídica sobre la materia ha de partir de la distinción entre aquellos casos en que el uso de los materiales protegidos por DPI en el entrenamiento de la IA tiene como finalidad extraer valor de los elementos expresivos de la obra de aquellos casos en los que no lo tiene. O, dicho de otro modo, ha de diferenciarse entre aquellos casos en que la copia tiene como finalidad acceder a la parte de la obra que la propiedad intelectual no protege (las ideas, los hechos o la información) de aquellos casos en que la copia tiene como finalidad acceder a lo que la propiedad intelectual sí que protege (la expresión de esas ideas, hechos o información). Esta distinción es especialmente importante en aquellas jurisdicciones que tienden a dispensar un mayor grado de protección a los titulares de DPI, como la UE, pues permitirá dar con soluciones que pondren adecuadamente los intereses en juego y no restrinjan innecesariamente la innovación en casos donde los intereses de los titulares de los DPI no se ven afectados en modo alguno.

En el plano comparado, existe cierto consenso en torno a la tesis de que el entrenamiento de IA con obras y prestaciones objeto de propiedad intelectual debe permitirse, sin consentimiento de los titulares y sin compensación económica, cuando el propósito del uso es que los sistemas de IA «aprendan» de los elementos no protegidos por la propiedad intelectual (las ideas). En este sentido se ha pronunciado el *case law* en EEUU y la legislación japonesa, así como un sector doctrinal muy autorizado de la doctrina europea. Las excepciones a la MTD introducidas en la DDAMUD, sin embargo, no parten de esta premisa, aunque sería esperable una revisión de la regulación en los próximos años por las razones apuntadas.

Mayores dudas plantea el aprendizaje de la IA a partir de obras y prestaciones objeto de propiedad intelectual cuando lo que se busca es que los sistemas de IA extraigan valor de los elementos expresivos, que es precisamente lo que la propiedad intelectual protege. Si bien la legislación japonesa ha optado decididamente por permitir también con carácter general este supuesto de entrenamiento de IA sin autorización de los titulares y sin compensación económica, no está claro en EEUU si la doctrina del *fair use* amparará estos usos ni en la UE cómo se abordará legislativamente la cuestión. Tampoco parece claro, a la vista de las distintas visiones sobre la materia, que se llegue a una solución similar en las distintas jurisdicciones, como sí es esperable que ocurra en el caso del entrenamiento de IA con obras y prestaciones con el propósito de acceder a los elementos no protegidos por la propiedad intelectual.

Ante las muchas incertidumbres existentes todavía en relación con la materia estudiada, hay una cuestión que está meridianamente clara: en el marco de la carrera mundial hacia el liderazgo en IA, la respuesta que se dé en cada jurisdicción a la pregunta que da título a este artículo tendrá mucho que ver en el resultado.

VII. BIBLIOGRAFÍA

- R. BERCOVITZ RODRÍGUEZ-CANO (Coord.), *Manual de Propiedad Intelectual*, 9^a edición, Tirant Lo Blanch, Valencia, 2019.
- M. W. CARROLL, «Copyright and the Progress of Science: Why Text and Data Mining is Lawful», *Washington College of Law Research Paper No. 2020-15*.
- M. CASPERS, L. GUIBAULT, «Baseline report of policies and barriers of TDM in Europe», *Future TDM*, septiembre 2015. https://www.futuretdm.eu/wp-content/uploads/FutureTDM_D3.3-Baseline-Report-of-Policies-and-Barriers-of-TDM-in-Europe-1.pdf
- T. CHIOU, «Copyright Law and Algorithmic Creativity: Monopolizing Inspiration?», *EU Internet Law in the Digital Single Market* (Ed. T. Synodinou, P. Jougleux, C. Markou, T. Prastituou-Merdi), Springer, Cham, 2021
- B. DELIPETREV, C. TSINARAKLI, U. KOSTIC, «Historical Evolution of Artificial Intelligence», Oficina de Publicaciones de la Unión Europea, Luxemburgo, 2020.
- J. DREXL, R. M. HILTY, L. DESAUNETTES-BARBERO, J. GLOBOCNIK, B. GONZÁLEZ OTERO, J. HOFFMAN, D. KIM, S. KULHARI, H. RICHTER, S. SCHEUERER, P. R. SLOWINSKI, K. WIEDEMANN, «Artificial Intelligence and Intellectual Property Law Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate», *Max Planck Institute for Innovation and Competition Research Paper No 21-10*, abril 2021.
- J. DREXL, R. M. HILTY, F. BENEKE, L. DESAUNETTES, M. FICNK, J. GLOBOCNIK, B. GONZÁLEZ OTERO, J. HOFFMAN, L. HOLLANDER, D. KIM, H. RICHTER, S. SCHEUERER, P. R. SLOWINSKI, J. THONEMANN, «Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective», *Max Planck Institute for Innovation & Competition Research Paper No. 19-13*, octubre 2019.
- R. DUCATO, A. STROWEL, «Limitations To Text and Data Mining and Consumer Empowerment. Making the Case for a Right to «Machine Legibility»», *CRIDES Working Paper Series*, 31 de octubre de 2018.
- M. DUQUE LIZARRALDE, «Las obras generadas por inteligencia artificial, un nuevo reto para la propiedad intelectual», *Pe. i. revista de propiedad intelectual*, nº 64, enero-abril 2020.
- P. FERNÁNDEZ CARBALLO-CALERO, *La propiedad intelectual de las obras creadas por inteligencia artificial*, Aranzadi, Cizur-Menor (Navarra), 2021.
- S. FLYN, C. GEIGER, J. QUINTAIS, T. MARGONI, M. SAG, L. GUIBAULT, M. CARROLL, «Implementing User Rights for Research in the Field of Artificial Intelligence», *Washington College of Law Research Paper No. 2020-12*.
- A. GARCÍA VIDAL, «Propiedad intelectual y minería de textos: Estudio de los artículos 3 y 4 de la Directiva (UE) 2019/790», *Actas de Derecho Industrial 40*, 2019-2020.
- C. GEIGER, «The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and The Failed Text and Data Mining Limitations in the CSDM Directive», *American University Washington College of Law*, PIJIP/TLS Research Paper, 66, abril 2021.
- C. GEIGER, G. FROSIO y O. BULAYENKO, «Text and Data Mining: Articles 3 y 4 of the Directive 2019/790/EU», *Propiedad intelectual y mercado único digital europeo* (Dir. C. Saiz García y R. Evangelio Llorca), Tirant Lo Blanch, Valencia, 2019.
- B. GONZÁLEZ OTERO, «Las excepciones de minería de textos y datos más allá de los derechos de autor: la ordenación privada contraataca», *Propiedad intelectual y*

- mercado único digital europeo (Dir. C. Saiz García y R. Evangelio Llorca), Tirant Lo Blanch, Valencia, 2019.
- J. GRIMMELMAN, «Copyright for Literate Robots», *Iowa Law Review*, 101:657, 2016.
- HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, «A definition of AI: Main capabilities and disciplines», Bruselas, abril 2019.
- R. HILTY y H. RICHTER, «Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules Part B Exceptions and Limitations (Art. 3 – Text and Data Mining)», *MPI Innovation and Competition Research Paper No. 17-02*, 2019.
- M. KOP, «The Right to Process Data for Machine Learning Purposes in the EU», *Harvard Journal of Law & Technology*, volumen 34, *Digest Spring 2021*.
- D. LEHR y P. OHM, «Playing with the Data: Why Legal Scholars Should Learn About Machine Learning», *University of California Davis Law Review*, 51, 2017, pp. 655-717
- M. A. LEMLEY, B. CASEY, «Fair Learning», *Texas Law Review*, volumen 99, nº 4, 2021.
- T. MARGONI, «Artificial Intelligence, Machine Learning and EU copyright law: Who owns AI?», *CREATE Working Paper 2018/12*, diciembre 2018.
- T. MARGONI, M. KRETSCHMER, «A deeper look into the EU Text and Data Mining exceptions: Harmonisations, data ownership, and the future of technology», *CREATE Working Paper 2021/7*.
- T. MARGONI, M. KRETSCHMER, «The Text and Data Mining exception in the Proposal for a Directive on Copyright in the Digital Single Market: Why it is not what EU copyright law needs», 25 abril 2018.
- E. ROSATI, «Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity», *Asia Pacific Law Review*, 26 de septiembre de 2019.
- M. SAG, «The New Legal Landscape for Text Mining and Machine Learning», *Journal of the Copyright Society of the USA*, Vol. 66, febrero 2020.
- S. SAMOILI, M. LÓPEZ COBO, E. GÓMEZ, G. DE PRATO, F. MARTÍNEZ-PLUMED y B. DELIPETREV, «AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence», Oficina de Publicaciones de la Unión Europea, Luxemburgo, 2020.
- R. SÁNCHEZ ARISTI, «La sentencia de apelación en el caso Google Books y los límites del fair use: una oportunidad para los autores y editores europeos», *La Ley*, nº 8654, 26 de noviembre de 2015.
- K. SCHWAB, *La cuarta revolución industrial*, Debate, Barcelona, 2016.
- M. SENFTLEBEN, T. MARGONI, D. ANTAL, B. BODO, S. VAN GOMPEL, C. HANDKE, M. KRETSCHMER, J. POORT, J. QUINTAIS, S. SCHWEMER, *Ensuring the Visibility and Accessibility of European Creative Content on the World Market: The Need for Copyright Data Improvement in the Light of New Technologies*, 2021. <https://ssrn.com/abstract=3785272>
- B. L. W. SOBEL, «Artificial Intelligence's fair use crisis», *Columbia Journal of Law & The Arts*, 41, 2017.
- B. SOBEL, «A Taxonomy of Training Data: Disentangling the Mismatched Rights, Remedies, and Rationales for Restricting Machine Learning», *Artificial Intelligence and Intellectual Property*, (Coord. J. Lee, R. Hilty, K. Liu), Oxford University Press, Oxford, 2021.
- T. UENO, «The Flexible Copyright Exception for 'Non-Enjoyment' Purposes - Recent Amendment in Japan and Its Implication», *GRUR International*, volumen 70, número 2, febrero 2021.