



Nuevos análisis en torno a los derechos de autoría e intangibles (artículos de investigación).

La “reutilización justa de datos”: una posible regulación para el entrenamiento de datos en herramientas de IA generativas.

Sin duda estamos ante una nueva revolución tecnológica, y en este momento en que estamos aún adaptándonos legalmente a las herramientas de IA, nos encontramos ante uno de los temas que más debates y más análisis está suscitando a nivel global, también desde el punto de vista jurídico. Las herramientas de IA generativa plantean problemas con los derechos de autoría que ostentan personas autoras y empresas titulares de todo tipo de contenidos publicados en internet, contenidos que son utilizados por las empresas desarrolladoras de estas tecnologías como datos de entrenamiento y de aprendizaje automático. Este uso plantea interrogantes, y ya ha propiciado las primeras demandas judiciales en distintos territorios, bajo la alegación de una presunta vulneración de los derechos exclusivos de autoría. Los titulares de derechos plantean la firma de acuerdos de licencia para permitir esta nueva explotación de sus contenidos. Las empresas desarrolladoras de IA por su parte se acogen al sistema regulado de excepciones y límites para justificar dicha explotación. Uno de los conceptos que ha aparecido recientemente y que podría ser de aplicación a esta problemática legal es la de la “reutilización justa” de los datos. En este artículo se analiza qué es la “reutilización” y si puede ser la tan ansiada solución legal a este conflicto.

© Conchi Cagide Torres. Asociación Intangia. Navarra. 2024. ISNI: 0000000506286844

El **derecho de autoría** consiste en una serie de facultades morales y patrimoniales exclusivas, aunque también hay una serie de **límites o excepciones** a los derechos, que permiten la explotación de los contenidos intelectuales sin autorización de la parte autora. En territorio Europeo, las excepciones se aplican bajo 3 condiciones concretas (los llamados 3 pasos):

Estos límites deben aplicarse de forma específica a casos concretos

No pueden perjudicar la normal explotación de la obra intelectual

No pueden afectar de forma no justificada a los intereses legítimos del autor

En ámbitos territoriales de influencia anglosajona tienen sus propias condiciones que pueden ser también catalogadas como excepciones a los derechos exclusivos de autoría. Es el **fair use** o uso justo de contenidos protegidos, que se aplica bajo reglas como las siguientes:

El propósito y carácter del uso

La naturaleza de la obra protegida

La cantidad y sustancialidad del uso

El impacto en el mercado

Por último asistimos con cierta incertidumbre a la aplicación de otras posibilidades legales que también podrían ser de utilidad: la **reutilización justa de los datos**. Este no es un concepto nuevo.

En territorio Europeo ya se usa en las normativas que regulan los datos abiertos cuyo origen está en el ámbito público. Este artículo analiza si es posible también su aplicación al sector privado, en concreto, al ámbito de las herramientas de IA generativa.

La “reutilización justa de datos”: una posible regulación para el entrenamiento de datos en herramientas de IA generativas.

Los conceptos de límite o excepción, “uso justo”, “uso inocuo” y “reutilización” de datos

Es importante reconocer que los derechos de propiedad intelectual no son derechos absolutos. Determinados usos están permitidos, en aras del interés general o público, no podemos obviarlos, aunque la propia legislación ha creado reglas para compensar a las personas autoras, lo que sí nos permite obtener es, a cambio de esos usos inevitables, una compensación económica.

Los **límites** a los derechos de propiedad intelectual tienen un largo recorrido, ya el Convenio de Berna de 1886 los regula en su art. II; en territorio Europeo, se han ido trasladando y adaptando a la regulación sobre derechos de autor, por ejemplo, la más reciente la Directiva 2019/790/UE sobre derechos de autor en el mercado único digital, que los amplía considerablemente, para recoger, por primera vez, un límite como el de la **minería de textos y datos**. El Considerando 8 de la Directiva indica que *“las nuevas tecnologías hacen posible el análisis computacional automatizado de información en formato digital, por ejemplo de textos, sonidos, imágenes o datos, al que generalmente se denomina «minería de textos y datos». La minería de textos y datos posibilita el tratamiento de grandes cantidades de información con el fin de adquirir nuevos conocimientos y descubrir nuevas tendencias”*.

Conforme a los art. 66 y 67 del **Real Decreto Ley 24/2021, de 2 de noviembre**, que traspone dicha Directiva y modifica la Ley de Propiedad Intelectual española, la minería de textos y datos es toda técnica analítica automatizada destinada a analizar textos y datos en formato digital a fin de generar información que incluye pautas, tendencias, correlaciones u elementos similares. Por lo tanto, dicho límite podría aplicarse a herramientas de IA y a su entrenamiento con datos de cualquier tipo, incluyendo también las herramientas de IA generativa.

No olvidemos que los límites que se aplican a los derechos de propiedad intelectual deben cumplir con la regla de los 3 pasos que aparece en el art. 40 bis de la LPI ya que deben aplicarse:

- ⇒ **Solo a casos concretos**; este paso se cumple ya que la excepción de minería de textos y datos se aplica al uso de datos por parte de organismos de investigación e instituciones responsables del patrimonio cultural y respecto de las obras y prestaciones intelectuales que gestionen o a las que den acceso, si disponen de una licencia previa de los titulares de derechos o por tratarse de obras en dominio público; igualmente se aplica sobre datos o contenidos sustanciales que se contienen en una base de datos que ya ha sido divulgada; además el límite no se aplicará si el autor o titular ha insertado una licencia que prohíbe el uso de los datos para estos fines de minería de textos y datos;
- ⇒ **Siempre que no perjudiquen la normal explotación de la obra intelectual**: este paso es el más complejo en cuanto a su prueba, y se refiere tanto a la obtención de un beneficio económico con el proceso de la minería de textos y datos, como al posible perjuicio en el mercado de la obra intelectual que es utilizada sin autorización; cuanto más específico sea ese uso, más afecta al mercado en cuestión;
- ⇒ **Siempre que no perjudiquen de forma injustificada los intereses de los autores**: el tercer paso hace referencia tanto a los derechos morales como patrimoniales del autor, y normalmente se salva si al límite se aplica un derecho de remuneración compensatoria; sobre esta posibilidad, para los procesos de minería de textos y datos que realicen en el entorno de instituciones de patrimonio cultural o centros de investigación, la Directiva 790/2019/UE no ha previsto aplicar un régimen de remuneración compensatoria; y para otros casos en los que se pueda aplicar el límite, se ha optado por una solución distinta, reflejada en el Considerando 18 de la Directiva, que es que los autores y titulares de derechos puedan aplicar reservas o prohibiciones a la aplicación del límite. Se sustituye aquí la compensación económica por una condición contractual individual.

La “reutilización justa de datos”: una posible regulación para el entrenamiento de datos en herramientas de IA generativas.

Los conceptos de límite o excepción, “uso justo”, “uso inocuo” y “reutilización” de datos (continuación)

En algunos territorios se aplican reglas distintas a la de los límites o excepciones y el sistema de los 3 pasos que se regula en el Convenio de Berna. En algunos estados, como por ejemplo, Estados Unidos de América, así como otros de corte anglosajón, el sistema que se utiliza para permitir el uso de contenidos sin autorización es el denominado **fair use o uso justo**. Podemos preguntarnos igualmente si esta norma del uso justo se puede aplicar a contenidos intelectuales que se utilizan para el entrenamiento automatizado de los sistemas de IA.

Como en el caso de los límites o excepciones aplicables en Europa, en el caso del *fair use* anglosajón, la respuesta debe ser precedida por el análisis de 4 factores concretos

- ⇒ **El propósito y carácter del uso:** se valora si el uso de contenidos ajenos se realiza con fines comerciales o no comerciales, y si por lo tanto el carácter de dicho uso es lucrativo o no lucrativo; en el caso en que sea no lucrativo, entrará dentro del *fair use* o uso justo. Además, el propósito del uso se refiere al requisito de ser un uso transformador o para propósitos diferentes; cuanto más transformador sea el uso o cuanto más diferente sea el propósito, más uso justo, si fuera una copia exacta para el mismo propósito, no podría entrar dentro del uso justo. Este requisito deberá ser analizado para los supuestos de uso de datos y contenidos para el entrenamiento de sistemas de IA, y también para sistemas de IA generativa; por ejemplo, el carácter no lucrativo no se cumple si las empresas desarrolladoras de IA obtienen beneficios económicos de los usuarios finales con las licencias o bien obtienen beneficios económicos de otra forma (con la publicidad, por ejemplo); en cuanto al uso transformador, este se produce, como hemos señalado, si se modifica una obra original o bien esta se utiliza para propósitos o fines diversos; el uso del código de un software de una empresa competidora para comprender los elementos funcionales del mismo o el uso para una nueva utilidad, que va a contribuir al aumento del conocimiento sobre el software, se consideran un uso transformador. En el caso de herramientas de IA, si el aprendizaje automático o el entrenamiento con los datos es una nueva funcionalidad o utilidad, se considera uso transformador, y por lo tanto, puede acogerse al uso justo. Sin embargo, si el uso de contenidos intelectuales con estos fines de entrenamiento cumple un propósito idéntico al contenido original (propósito estético, expresivo, funcional, etc) no se cumplirá ese efecto transformador.
- ⇒ **La naturaleza de la obra protegida:** en este factor se valora si el contenido utilizado para el entrenamiento de IA es un contenido inédito, o por el contrario, ya ha sido publicado, en el primer caso, no entrará dentro del uso justo. También se analiza si el contenido es creativo o son meros datos; en el primer caso, no entraría dentro del uso justo, sí entraría si se usan meros datos para el entrenamiento de los sistemas de IA.
- ⇒ **La cantidad y sustancialidad del uso:** en este factor se valora no solo si el contenido que se utiliza es solo un fragmento o se usa completo, sino también, en el primer caso, si ese fragmento tiene importancia en relación al conjunto del contenido; se analiza si el uso de un contenido completo es razonable, esto sucede con las herramientas de IA, normalmente se utilizan contenidos completos, lo requiere para poder aprender los patrones específicos, entonces se analizará la sustancialidad del uso, si ese uso es razonable en relación al propósito, que en el caso de sistemas de IA, está relacionado con el aprendizaje de patrones inherentes a los medios generados por humanos. Si parte del contenido es copiado de forma literal, no se aplicaría un uso justo.
- ⇒ **El impacto en el mercado:** si el uso justo de un contenido busca competir en el mismo mercado que la obra original, sustituyéndola en dicho mercado, no estaremos ante un uso justo; en el caso de herramientas de IA, pueden competir con los contenidos humanos, pero deberá analizarse si ese impacto es positivo o negativo.

La “reutilización justa de datos”: una posible regulación para el entrenamiento de datos en herramientas de IA generativas.

Los conceptos de límite o excepción, “uso justo”, “uso inocuo” y “reutilización” de datos (continuación)

Además de las reglas aplicables a límites o excepciones a los derechos de autor, o de los factores que permiten aplicar el sistema de uso justo o *fair use* en territorios anglosajones, no podemos olvidar de traer a esta investigación las decisiones jurisprudenciales que hayan podido dictaminarse sobre esta materia, analizando si tendrían acogida a sistemas como las herramientas de IA, incluyendo también a las de IA generativa.

Una de estas decisiones tiene un gran calado en esta materia, ya que plantea una tercera posibilidad, la del **uso inocuo del derecho ajeno (*ius usus inocui*)**. Este concepto no es nuevo ni mucho menos, se recoge en el art. 7 del Código Civil que regula el ejercicio de los derechos conforme a las exigencias de la buena fe y bajo la prohibición del abuso del derecho o el ejercicio antisocial del mismo. La sentencia española concreta que lo aplica a los límites de los derechos de autoría es la Sentencia del Tribunal Supremo (TS) nº 172/2012, de 3 de abril (Rec nº 2037/2008), en un asunto en el que el titular de una página web demanda a Google por copiar parte del código htm de su página web, en la memoria caché del buscador y en el sistema de búsqueda de contenidos en internet. El TS resuelve el recurso sobre la Sentencia de la Audiencia Provincial de Barcelona que había aplicado el *fair use* norteamericano en vez del art. 40 bis LPI o la regla de los 3 pasos.

En la sentencia recurrida de la AP de Barcelona se mencionaba que la interpretación correcta del artículo 40 bis es la doctrina del “fair use”, lo que fue empleado por el recurrente para alegar que se había aplicado una doctrina ajena al derecho español para interpretar una norma de nuestro ordenamiento jurídico (ESTEVE, MC BRIDE, MARTÍNEZ Y POYSKY, Pg.3)

Por lo tanto, es una decisión en este caso trascendental para el tema de este artículo ya que alude a la compatibilidad entre unos y otros sistemas que limitan los derechos exclusivos de autoría. Declara el TS que la regla de los 3 pasos es una manifestación del **uso inocuo del derecho ajeno** o principio general del derecho del *ius usus inocui*. Bajo esta regla, ya aplicada en otras decisiones jurisprudenciales precedentes, y que sigue utilizándose para el sector de los derechos de propiedad intelectual (por ejemplo, se ha utilizado por los Juzgados en 1ª instancia en el famoso caso de la demanda entre Vegap y Mango, en relación a obras plásticas que se han transformado en NFT,s, CENDOJ (Roj: SJM B 1/2024 – ECLI:ES:JMB:2024:1), se entiende que el uso de un derecho ajeno es inocuo, si ese uso favorece a su titular y no le perjudica; en el caso que estamos analizando de Google, parece evidente a ojos del TS que la copia de una parte de la web del demandante en el motor de búsqueda no perjudica la normal explotación de dicha obra intelectual, al contrario, le beneficia (permite que sea fácilmente localizable y así aumenta la difusión de la misma).

Bajo este criterio, se analizaría si el uso de contenidos por herramientas de IA generativa se puede contemplar como uso inocuo, evaluando **si ese uso es beneficioso para esos contenidos** o si ese uso **contribuye a una mejor difusión** de los mismos sin perjudicar su normal explotación.

En aplicación del principio de especialidad conforme al cual la norma especial que es la LPI y que es de aplicación preferente a la norma general del Código Civil, creo que debe aplicarse antes la regla de los 3 pasos, pero debo reconocer que las decisiones jurisprudenciales sobre la materia sientan precedente y tienen su importancia como vías interpretativas en futuros casos relacionados con la IA generativa. Por otro lado, sería necesario completar este análisis con el límite recogido en el art. 31.1. de la LPI, de reproducciones provisionales o transitorias que forman parte de un proceso tecnológico; se trata de una excepción obligatoria, aplicable a las copias efímeras o intermedias, que podría dar lugar a un nuevo artículo de investigación, para analizar el concepto de uso efímero de las obras intelectuales y su aplicabilidad al uso de contenidos ajenos para el entrenamiento de sistemas de IA.

La “reutilización justa de datos”: una posible regulación para el entrenamiento de datos en herramientas de IA generativas.

Los conceptos de límite o excepción, “uso justo”, “uso inocuo” y “reutilización” de datos (continuación)

Y así llegamos al concepto de “reutilización” de los datos. Un concepto que se ha debatido recientemente y en profundidad en territorio Europeo, en un ámbito concreto, el del uso de datos abiertos y su reutilización en el sector público. La Directiva 2019/1024/UE, denominada comúnmente la Directiva sobre datos abiertos, regula la **reutilización de los datos en el sector público**. Posterior a esta normativa, el Reglamento Europeo 2022/868 sobre gobernanza de datos, y su Reglamento de ejecución 2023/2138, completan la regulación sobre la materia, introduciendo un concepto nuevo, el de la “**gobernanza algorítmica**”, que analizamos a continuación.

Antes de estas regulaciones europeas, ya se había incorporado el concepto de reutilización vinculado a las nuevas tecnologías en la regulación de las **bases de datos electrónicas** con la Directiva 1996/9/CE. En dicha normativa se regula el **derecho de reutilización de las bases de datos** como “*toda forma de puesta a disposición del público de la totalidad o de una parte sustancial del contenido de la base mediante la distribución de copias, alquiler, transmisión en línea o en otras formas. La primera venta de una copia de una base de datos en la Comunidad por el titular de los derechos o con su consentimiento extinguirá el derecho de control de las ventas sucesivas de dicha copia en la Comunidad*” (art. 7.2. b) de la Directiva).

El Tribunal de Justicia Europeo en STJUE sobre el asunto C-202/12, asunto Innoweb, define la reutilización como el acceso a todo el contenido de una base de datos por un **medio distinto** del previsto por el fabricante de dicha base de datos. Por su parte, en el asunto C-762/19, asunto CV online, el TJUE al resolver sobre la reutilización de una base de datos de acceso libre en internet, establece el requisito de que la reutilización debe referirse a una parte sustancial del contenido reutilizado, y se vulnera ese derecho de reutilización si **se pone a disposición del público esa parte sustancial** (o la base de datos completa) si se hace sin consentimiento del fabricante de la base de datos y esta puesta a disposición priva a aquel de los ingresos derivados del negocio con la base de datos.

Cabría plantearse si algunos de los contenidos utilizados para el entrenamiento de las herramientas de IA son bases de datos electrónicas (los artículos del New York Times, o el banco de imágenes de Gettyimages, por ejemplo; estos contenidos utilizados por herramientas de IA han sido objeto de demandas judiciales contra OpenAI).

Sin embargo, con las nuevas normativas Europeas sobre datos abiertos, se pasa de considerar el derecho de reutilización como un derecho exclusivo a regular principios como el de que los datos públicos o datos recogidos con fondos públicos deben ser reutilizables para fines diversos, sean estos comerciales o no comerciales.

El artículo 2 apartado 11 de la Directiva 1024/2019/UE de datos abiertos, define la “reutilización” como:

“El uso por personas físicas o jurídicas de documentos que obran en poder de:

- a) organismos del sector público, con fines comerciales o no comerciales distintos del propósito inicial que tenían esos documentos en la misión de servicio público para la que se produjeron, excepto para el intercambio de documentos entre organismos del sector público en el marco de sus actividades de servicio público, o*
- b) empresas públicas, con fines comerciales o no comerciales distintos del propósito inicial que tenían esos documentos de prestar servicios de interés general para el que se produjeron, excepto para el intercambio de documentos entre empresas públicas y organismos del sector público que se realice exclusivamente en el desarrollo de las actividades de servicio público de los organismos del sector público”.*

¿Sería posible aplicar esta regulación a herramientas de IA generativa? Veremos a continuación en qué afecta esta regulación y qué relación tiene con la nueva y más reciente normativa sobre gobernanza de datos, el Reglamento 2022/868.

La “reutilización justa de datos”: una posible regulación para el entrenamiento de datos en herramientas de IA generativas.

Las normativas Europeas reguladoras de la “reutilización” de datos abiertos: la Directiva 1024/2019/UE

La Directiva 1024/2019 tiene una función específica de reutilización de datos creados o financiados por el sector público, y también tiene como finalidad la adaptación de las funciones públicas a la era de las nuevas tecnologías. De hecho, el Considerando 13 de dicha Directiva reconoce que la información producida, reproducida, recogida o difundida en ejercicio de una función pública puede ser un **recurso importante para el desarrollo de tecnologías avanzadas**, citando expresamente a los sistemas de IA y al internet de las cosas.

Aunque esta reutilización debe realizarse sobre datos y bajo criterios o condiciones reguladas en la propia Directiva:

⇒ En primer lugar, la normativa se refiere al tipo de dato público o dato abierto: es aquel tipo de dato que se puede utilizar, reutilizar y compartir libremente, para cualquier fin, si cumplen con ese elemento de ser “**datos que obran en poder de organismos públicos**” (Considerando 14 de la Directiva).

La pregunta que podemos hacernos es si se refieren también a datos que son de titularidad de personas físicas o entidades no públicas, pero que, por distintos motivos, han llegado a estar en poder de organismos públicos.

⇒ En segundo lugar, la reutilización se refiere a usos de esos datos (documentos) para **finés distintos** a los que motivaron la generación, uso o difusión del mismo; ese fin distinto puede ser cualquier actividad, incluida la comercial, ya que la Directiva alude a usos que los organismos públicos permitan a través de licencias, o venta (Considerando 22).

Otra pregunta es si ese uso comercial para fines distintos estará limitado al propio organismo público que es poseedor del dato o el uso comercial se permite a otros organismos, públicos o privados.

⇒ En tercer lugar, la reutilización se aplica a **todo tipo de datos o documentos**, indicando expresamente que abarca todas las formas de representación de actos, hechos o información, y su recopilación, con independencia del formato o soporte (papel, audiovisual, grabación sonora, etc); conforme al Considerando 30, parece no incluir los programas informáticos; sí hace una referencia específica a datos derivados de la investigación científica, como los artículos científicos, sin más restricciones que el reconocimiento de la autoría; aunque también reconoce la propia Directiva que se tendrán en cuenta aspectos como la protección de datos personales, la confidencialidad, los secretos comerciales o los **derechos de propiedad intelectual de terceros**, entre otros derechos aplicables.

No es lo mismo la reutilización de un dato estadístico, como un dato meteorológico o económico, que un dato que está relacionado con obras y contenidos intelectuales, o un documento que contenga una obra protegida por la LPI.

En este sector cultural en concreto, reconoce la Directiva que los datos de organismos públicos como museos, bibliotecas y archivos pueden requerir, para su reutilización, inversiones económicas relacionadas con esta finalidad de reutilización de documentos que están protegidos por la propiedad intelectual. Y que dichas inversiones económicas pueden requerir de cooperación con socios privados, que quieran de alguna forma rentabilizar dicha inversión con derechos exclusivos. Sin embargo, esta Directiva, tal y como refleja el artículo 1.5, debe ser compatible con las **normativas que protegen los derechos de propiedad intelectual**; no pueden quedar afectados los derechos de propiedad intelectual de terceros (Considerando 54). Tampoco se puede hablar de reutilización si los datos o documentos se refieren a patentes, marcas o diseños industriales registrados.

La “reutilización justa de datos”: una posible regulación para el entrenamiento de datos en herramientas de IA generativas.

Las normativas Europeas reguladoras de la “reutilización” de datos abiertos: la Directiva 1024/2019/UE (continuación)

Sin embargo, sí que indica que si los titulares de derechos de propiedad intelectual son los mismos organismos públicos estos no podrán ejercer esos derechos en contra de la reutilización. Reconoce los **derechos morales** del personal integrado en dichas entidades públicas, pero también el derecho de explotación en beneficio de la entidad y que debe **prevalecer la reutilización** y no el uso exclusivo. Por ejemplo, si un museo o entidad cultural tiene un acuerdo con un agente privado para la digitalización en exclusiva de los recursos culturales que gestiona, dicho acuerdo no debería tener una vigencia superior a 10 años.

⇒ Uno de los apartados que más sorprenden es la posibilidad de **conceder licencias** (incluso exclusivas) para cumplir con la finalidad de reutilización de los datos o documentos.

En relación a los documentos y fondos relacionados con el patrimonio cultural, reconoce que su reutilización puede ser especialmente productiva en sectores como el turismo, o el educativo; otras instituciones culturales, y cita expresamente a **orquestas, óperas, ballets y teatros**, por su área de especialidad, deben ser excluidas de la aplicación de esta Directiva, debido a que prácticamente la totalidad de esos documentos son material protegido por derechos de titularidad de terceros.

En el caso de las licencias, la Directiva recomienda el uso de **licencias tipo**, que estén a disposición públicamente para cumplir con el principio de transparencia, y preferentemente, se trate de **licencias abiertas** (art. 8 y 14 de la Directiva).

Por licencias abiertas entendemos que sean licencias Creative Commons, aunque estará por ver qué tipo de condiciones se eligen; recordemos condiciones como el reconocimiento, obra no derivada, uso no comercial o compartir igual; y también se analizará si esas condiciones casan bien con algunos límites de esta Directiva, como la duración o la exclusividad.

La Directiva 1024 ya ha sido integrada en el ordenamiento jurídico español, a través del Real Decreto-Ley 24/2021, de 2 de noviembre. En la misma línea, la norma española excluye de la reutilización de los datos a aquellos documentos sobre los que existen derechos de propiedad intelectual vigentes, aunque también indica que si esos documentos están en poder de museos, bibliotecas o archivos de titularidad pública, o se trata de datos de investigación (financiados públicamente) se aplicará la reutilización si estas instituciones culturales, universidades o centros de investigación (públicos o financiados con fondos públicos, insistimos en este detalle) sean titulares de los derechos, debiendo respetar siempre los términos de la cesión de derechos contemplada en la LPI. Igualmente, se indica que si son fabricantes de bases de datos, los organismos públicos no restringirán ni la extracción ni la reutilización de dichas bases de datos (art. 4.4. del Real Decreto-Ley).

De una lectura pormenorizada de la Directiva sobre datos abiertos y reutilización del sector público podemos concluir que la reutilización puede tener como consecuencia que los datos y documentos que tienen las entidades públicas en su poder puedan ser reutilizados para herramientas de IA generativa, pero debe cumplirse con la normativa de propiedad intelectual vigente, de forma que deba proporcionarse esta información siempre bajo licencias, y que dichas licencias deben ser transparentes. Me preocupa más que estas reglas de la LPI no se apliquen si los autores son personal empleado de las propias entidades públicas.

Analizamos a continuación si los Reglamentos Europeos posteriores, sobre gobernanza de datos, normas Europeas de aplicación directa que no necesitan trasposición a nuestra legislación, han cambiado en algo esta regulación.

La "reutilización justa de datos": una posible regulación para el entrenamiento de datos en herramientas de IA generativas.

Las normativas Europeas reguladoras de la "reutilización" de datos abiertos: el Reglamento 2022/868 de gobernanza de datos.

Las instituciones Europeas, en aras del objetivo principal de facilitar la disponibilidad de datos y su reutilización en entornos confiables y también comerciales, para su uso en ámbitos de investigación e innovación tecnológica, aprueban el llamado **Reglamento de gobernanza de datos** (Reglamento UE 2022/868, de 30 de mayo) Y recalcamos la finalidad comercial, ya que uno de los apartados de la Estrategia Europea para los datos es precisamente, impulsar la economía de dichos datos; el modelo de negocio se basa no solo en poner a disposición los datos públicos y privados para su reutilización, sino también en los sistemas que completan dicha reutilización, como el punto de acceso único de datos a nivel Europeo, los servicios comerciales de intermediación de los datos y su registro o el Consejo Europeo de Innovación de datos, que se crean con este Reglamento y se desarrollan en el Reglamento de ejecución 2023/138 para los datos de alto valor. También se regula la finalidad altruista para datos de interés general que son objeto de cesión y la creación de un registro público para las organizaciones que se acojan a esta finalidad no comercial.

El Reglamento de gobernanza de datos por primera vez regula el uso y reutilización de los **datos que no son personales**, recogiendo en este concepto a aquellos que estén protegidos como secretos comerciales, **derechos de propiedad intelectual** y derechos de propiedad industrial, entre otros. Indica que en la reutilización de este tipo de datos no pueden quedar afectados los derechos de propiedad intelectual de terceros, y que en el caso de que así sea, deben aplicarse todas las condiciones (de protección, reconocimiento, cesión, etc) que se indican en las leyes vigentes de propiedad intelectual (Considerando 17). Para los datos distintos a estos, en poder de las Administraciones Públicas, ya se aplica la Directiva 2019/1024/UE, por lo que este Reglamento nº 868 de 2022 se aplica para datos que se excluyen de la Directiva anterior.

En la regulación relacionada con los servicios de intercambio de datos, sean estos servicios comerciales o no, excluye los servicios de intermediación de contenidos protegidos por los derechos de autor (por ejemplo, prestadores de servicios y contenidos digitales como redes sociales y plataformas de todo tipo). Unos y otros prestadores son diferentes: un servicio de intermediación de datos es un servicio cuyo objetivo es establecer relaciones comerciales para el intercambio de datos entre un nº indeterminado de interesados y titulares de los datos, por una parte, y usuarios por otra parte, a través de medios técnicos, jurídicos o de otro tipo (art.2.11 del Reglamento 868). El prestador intermediario de servicios y contenidos es aquel que ofrece servicios de acceso y navegación por internet, de transmisión de datos y de suministro de motores de búsqueda. El primero se regula en el art. 10 del Rgto. 868 y hace referencia al "mercado de datos" o intercambio de datos entre empresas y usuarios, a través de un servicio intermedio que debe cumplir con las condiciones recogidas en este Reglamento; entre las condiciones, la de que no pueden usar los datos (personales o no personales) que intercambian con fines lucrativos propios; un ejemplo es *DataSpace Europe*, servicio intermediario de datos económicos e industriales (Finlandia). El segundo se regula por el Reglamento Europeo 2022/2065 y un ejemplo de prestador intermediario de servicios y contenidos es el motor de búsqueda de Google. En los primeros, los intermediarios de los datos pueden firmar acuerdos de cesión o transmisión de dichos datos, bajo condiciones legales como la no exclusividad; si fuera necesario un acuerdo exclusivo, este acuerdo, conforme al Rgto. 868, no puede tener una vigencia superior a 12 meses. En los segundos, se aplican condiciones de uso generales a modo de contratos de adhesión con los usuarios del servicio, que tienen en muchos casos una duración indefinida.

Una de las cuestiones que planteamos es qué son las empresas desarrolladoras de sistemas de IA: ¿son intermediarios de datos, o son prestadores de servicios y contenidos digitales? Estos sistemas se usan por los propios prestadores de servicios digitales (siguiendo con el ejemplo de Google y su motor de búsqueda, se usan sistemas de IA para obtener resultados al hacer una búsqueda; Google sería un prestador de servicios solo o también un intermediario de datos?)

La "reutilización justa de datos": una posible regulación para el entrenamiento de datos en herramientas de IA generativas.

Gobernanza algorítmica, datos y herramientas de IA

Decía Daniel Innerarity en 2021 (ver en referencias bibliográficas) que *"gobernar es ya en gran medida —y lo será aún más— un acto algorítmico; una buena parte de las decisiones de gobierno son adoptadas por sistemas automatizados"*.

Cuando hablamos de **gobernanza algorítmica** hacemos referencia al conjunto de normas y prácticas aplicables a la creación y el uso de algoritmos integrados en sistemas de IA. Uno de los elementos normativos que se incluyen en dicha gobernanza algorítmica es el **Reglamento Europeo de IA** (que es pionero en la regulación de sistemas de IA a nivel mundial) que está en pleno proceso de aprobación. Pero también otras prácticas y normas anteriores a este Reglamento que adoptan la forma de códigos éticos de las empresas desarrolladoras, o que tienen que ver con las fases de recogida de los datos que van a alimentar estas herramientas. Por lo tanto, las dos normativas Europeas que hemos explicado en los apartados anteriores forman parte de la gobernanza algorítmica.

Así lo indican expresamente desde el Portal oficial de datos Europeos: los datos abiertos y la IA están estrechamente relacionados:

- ⇒ Por un lado, los datos abiertos pueden mejorar los sistemas de IA; si esta usa una gran cantidad de datos, los datos abiertos pueden reutilizarse para entrenar los sistemas de IA y así aumentar sus capacidades en forma de predicciones y razonamientos más precisos y más útiles;
- ⇒ Por otro lado, los sistemas de IA pueden analizar, con su capacidad analítica, esas grandes cantidades de datos que pueden ser reutilizados, identificando tendencias y patrones, de los que luego derivan grandes conocimientos.

Se puede apreciar cómo unos datos abiertos de alta calidad puede contribuir a unas herramientas de IA de alta calidad, y viceversa. Para todo ello, es imprescindible que esos datos sean de fácil acceso y puedan ser reutilizados para esos sistemas de alta capacidad tecnológica. Este es el objetivo de la **Directiva 1024/2019** analizada en apartados anteriores, que los datos públicos que están en poder de las entidades que realizan funciones públicas puedan ceder esos datos para su reutilización. Para estos datos públicos, distinguiremos entre datos de cualquier tipo que son creados en el seno de dichas entidades públicas, que sí son objeto de reutilización directa, de aquellos otros que pertenecen a terceros, con los que la entidad pública debe firmar el correspondiente acuerdo de cesión de derechos, de cara a dicha reutilización.

También se está planteando desde instancias administrativas el uso de los datos abiertos para los sistemas de **IA generativa**, que está utilizándose no solo para dar soluciones tecnológicas a cuestiones como la robótica, la industria, el diagnóstico médico, el cambio climático o la ciberseguridad, sino que también se ha desarrollado para generar creaciones intelectuales (artísticas, literarias, musicales, audiovisuales). Los sistemas de IA se alimentan de datos no personales, como obras y expresiones artísticas protegidas por las leyes de propiedad intelectual, para su entrenamiento. Estos materiales y contenidos entran dentro de la regulación del **Reglamento Europeo 868/2022**.

Confirmamos por lo tanto, que las normativas Europeas reguladoras de los datos abiertos y la gobernanza de datos permiten la reutilización de los mismos, para fines como el desarrollo y entrenamiento de sistemas de IA. En relación a datos no personales a los que se aplica la legislación de propiedad intelectual, ya que son contenidos protegidos con derechos vigentes, será fundamental analizar los distintos acuerdos que se firmen entre los cedentes de los datos no personales (Administraciones Públicas, empresas, etc) y los autores de dichos contenidos. Y por supuesto, analizar los posteriores acuerdos entre estos cedentes que permiten la reutilización de los datos no personales y los cesionarios de esos datos, así como analizar la labor que realicen los intermediarios de los datos en este escenario contractual.

La “reutilización justa de datos”: una posible regulación para el entrenamiento de datos en herramientas de IA generativas.

Conclusiones:

El uso de datos y contenidos para el entrenamiento y aprendizaje de herramientas de IA puede conseguirse bajo los paraguas de la legalidad y la legitimidad por varias vías:

- ⇒ A través de la aplicación de límites como la minería de textos y datos, si se cumplen con los 3 pasos del art. 40 bis de la LPI;
- ⇒ A través del *fair use* o uso justo anglosajón, siempre que se cumplan con los 4 factores obligatorios que lo permiten;
- ⇒ A través de la aplicación de reglas generales del derecho como el derecho de uso inocuo (Tribunal Supremo);
- ⇒ A través de las normativas que permiten la reutilización de los datos abiertos o datos que obran en poder de organismos públicos, especialmente aquellos datos o contenidos creados internamente por su personal;
- ⇒ A través de las normativas sobre gobernanza de datos, que permiten la reutilización de datos no personales, como contenidos protegidos por la legislación de propiedad intelectual, si los titulares de derechos así lo han permitido.

Las herramientas a disposición de las empresas desarrolladoras de IA son variadas, como podemos comprobar. Y una de las más potentes es sin duda, el sistema legal europeo de “reutilización justa” de los datos públicos y privados, en aras de la innovación tecnológica.

Nota final:

Durante los últimos 2 años se multiplican las noticias sobre demandas a OpenAI y a otras herramientas de IA generativa, en distintos territorios. Sigo con mucho interés el desarrollo de esas demandas, y los argumentos que publican tanto demandantes como demandadas, las empresas desarrolladoras de IA generativa; los primeros, alegan la vulneración de sus derechos de autor, las segundas quieren aplicar el *fair use* o uso justo. Algunos estados ya tienen regulación, como Japón, que tiene desde 2018 una regulación específica que permite la explotación de contenidos protegidos si se realiza para su uso en el análisis de datos (en este año 2024 va a publicar un informe para adaptar esta normativa). En Europa, mientras se concluye la aprobación del Reglamento de IA, se puede aplicar un límite a los derechos de autor basado en la minería de textos y datos (que ya fue aprobada en 2019). Y además, se permite reutilizar cualquier dato abierto y también datos no personales si estos se han creado en el seno de organismos públicos, o bien están en poder de dichos organismos públicos.

La posible solución son los acuerdos de licencia, es decir, la vía contractual: por un lado, esta vía puede servir a entidades de gestión colectiva para aplicar licencias (quieren que sean obligatorias) a las empresas desarrolladoras de IA; sus propuestas y manifiestos van en esta línea, aunque permitidme que dude de su eficacia, ya que estará por ver si estas entidades repartirán de forma “justa” las cantidades recaudadas a las personas autoras y a titulares de derechos. Por otro lado, también los contratos y licencias pueden servir a las entidades públicas para firmar con las empresas de IA, para la reutilización de todo tipo de datos con fines comerciales o no comerciales, siempre con fines de desarrollo e innovación tecnológica. Está una última vía, que ya está siendo utilizada por empresas como Meta, quienes son a la vez prestadoras de servicios digitales (con redes sociales como Facebook e Instagram) y empresas de IA, y que usan los contenidos sobre los que tienen derechos de explotación para el entrenamiento de sus sistemas de IA; como ya disponen de los derechos de explotación sobre todos los contenidos que los usuarios de esas redes sociales publicamos en las mismas, es decir, ya hemos firmado los correspondientes acuerdos de licencia, pueden hacer uso de los contenidos sin aplicar otras vías legales como límites, *fair use* o la reutilización de datos; sobre esta última vía, ni entidades de gestión colectiva ni organismos Europeos han tomado ningún tipo de medida más allá de exigir la necesaria transparencia, que, por cierto, sigue sin cumplirse en algunas cuestiones. Los usuarios lo único que podemos hacer, de momento, es oponernos a estos tratamientos. En ello estamos.

Conchi Cagide Torres.

La “reutilización justa de datos”: una posible regulación para el entrenamiento de datos en herramientas de IA generativas.



Bibliografía y otras referencias

Artículo “Resumen y comentario de la Sentencia Núm. 172/2012 del Tribunal Supremo, de 3 de abril de 2012, (MARIO contra GOOGLE-SPAIN S.L.)”. Lydia Esteve, Richard A. Mac Bride, Nuria Martínez y Athena Poysky. University of Alicante Intellectual Property and Information Technology. En www.uaipit.com [Fecha consulta: 24 mayo 2024].

Artículo “Datos abiertos e IA: una relación simbiótica para el progreso”. <https://data.europa.eu/en/publications/datastories/open-data-and-ai-symbiotic-relationship-progress>. [Fecha consulta: 13 junio 2023].

Artículo Innerarity, Daniel. “Gobernanza algorítmica”. 2021. <https://www.lavanguardia.com/opinion/20210731/7636808/gobernanza-algoritmica.html>

Sentencias del Tribunal de Justicia Europeo:

STJUE C-202/12 asunto INNOWEB

STJUE C-762/19 asunto CV-ONLINE

STS nº 172/2012, de 3 de abril (Rec nº 2037/2008),

Webs:

Blogs: <http://www.aladda.es>

Creative Commons: <https://www.creativecommons.org>

<https://data.europa.eu/es>

<https://observatorio-ametic.ai/en/aplicaciones-industriales-de-la-ia/estrategia-europea-de-datos>

Protección IA generativa en Japón:

<https://www.lexology.com/library/detail.aspx?q=d8b4ba7d-a764-4acd-adca-lb72a2f1aaa0>

Puedes acceder a todos los artículos de investigación en la web www.intangia.com

Autora: Conchi Cagide Torres. Directora del Departamento jurídico de Intangia..

Para citar a la autora:

© Conchi Cagide Torres. Asociación Intangia. Navarra. 2024. ISNI: 00000000506286844