ADAYA MARIA ESTEBAN RUIZ
Abogada

ENTRENAMIENTO DE REDES NEURONALES CON DATOS DE FUENTES PÚBLICAS, WEB SCRAPPING Y EL PROBLEMA DE LA FALTA DE TRANSPARENCIA Y TRAZABILIDAD DE LA INFORMACIÓN

1. Introducción

Son muchos los interrogantes que el uso de algoritmos y técnicas de inteligencia artificial nos plantean, sobre todo cuando queremos aplicar a la tecnología los populares (e incluso diría *manidos*) principios y derechos digitales[1]. Uno de ellos es la **transparencia y la libertad de elección y el respeto de la autonomía humana.**

En virtud del principio de transparencia, las personas deben ser informadas de cómo funciona el sistema de inteligencia artificial o el algoritmo, su trazabilidad, y explicabilidad, trazando el camino que ha recorrido y las razones que han llevado al sistema a tomar una decisión. Muy de la mano con el principio de transparencia, los últimos informes, guías y directrices hablan también de la necesidad de hacer que las tecnologías estén centradas en el ser humano y que las personas sean quienes mantengan, en todo momento, el poder y la libertad de elección.

A priori, parecería sencillo, pero… ¿cómo podemos asegurar la **transparencia y la libertad de elección de las personas** cuando utilizamos tecnologías para cuyo entrenamiento, a veces, nadie sabe muy bien de dónde viene la información? Hablamos desde el más simple de los algoritmos que permite el *web scrapping* hasta

[1] La última autoridad en pronunciarse ha sido la Comisión Europea que el pasado 26 de enero de 2022 propuso una serie de principios y derechos digitales, bajo la llamada Declaración europea de derechos y principios digitales (*European Declaration on Digital rights and principles*), centrados en el ser humano, que buscan servir de guía y de marco regulatorio para empresas y reguladores. Antes de la Comisión, el pasado 14 de julio, el Gobierno ya había presentado en España su particular Carta de Derechos Digitales, con la misma intención de servir de marco de referencia para los poderes públicos en relación con las potencialidades y retos del entorno digital y a su vez de declaración amplia de derechos que deberá ser, posteriormente, desarrollada por ley.

técnicas más avanzadas de inteligencia artificial, como las redes neuronales y el "aprendizaje profundo".

El uso de estas técnicas presenta problemas desde varias perspectivas desde la óptica legal, sabiendo en todo caso que la dificultad en estos casos reside en descubrir, tanto por parte de las autoridades como por los propios usuarios, qué se está usando esta técnica y que, gracias a ella, se está accediendo o recogiendo información, a veces catalogada como dato personal, a veces como secreto empresarial u otros. Esta dificultad es la que hoy en día anima, en gran medida, a su uso cada vez más masificado.

A continuación, haremos un repaso de los principales problemas que presentan desde distintas perspectivas, desde la protección de datos, hasta la propiedad intelectual y la incipiente regulación de la inteligencia artificial, los algoritmos y los derechos digitales.

2. Protección de datos: ¿Qué pasa cuando la información que se está utilizando para ese entrenamiento, proveniente de "fuentes públicas", tiene la categoría de "dato personal"?

2.1. Panorama actual en España.

A nivel de protección de datos, el mero "acceso" a datos personales entra dentro del ámbito de aplicación de la normativa de protección de datos y debe cumplir con las obligaciones ahí establecidas. Y ello, aunque esta información provenga de fuentes de acceso público o fuentes accesibles al público.

Como regulaciones que buscan establecer obligaciones de máximos, ni el Reglamento 2016/679, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos ("**RGPD**") ni la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales

("**LOPDGDD**") regulan el uso de este tipo de técnicas de *web scrapping* o de uso de información pública para entrenamiento de algoritmos, ni siquiera el acceso a información de "fuentes públicas".

Hasta el momento, tampoco las autoridades europeas y españolas en materia de protección de datos tampoco han dado su opinión o han publicado información relacionada con estas técnicas y solo la Agencia Española de Protección de Datos ("**AEPD**") ha realizado algún pronunciamiento en relación con la información proveniente de "fuentes públicas".

2.2. El tratamiento de datos personales de "fuentes públicas" es considerado un tratamiento de datos personales.

La mayor controversia se genera debido al mal uso del concepto de "fuentes accesibles al público" o información de "fuentes públicas". La idea principal que debemos mantener es que, aunque mediante el uso de estas técnicas, accedamos solo a datos disponibles en "*fuentes accesibles al público*", se considera que **existe un tratamiento de datos personales**. Los operadores del sector que utilizan estas técnicas se suelen apoyar en que el uso de estas técnicas, como el *web scrapping*, no implica el tratamiento de datos personales, dado que, al quedar estos datos disponibles al público general, todos pueden "hacer uso" de esta información y no necesitarían ninguna base legal para hacer tratamientos posteriores de estos datos personales.

En España, este error se deriva, en parte, de la anterior normativa existente en materia de protección de datos[2], que no se encuentra ya en vigor. Dicha normativa

[2] Nos referimos a la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal y su normativa de desarrollo, principalmente el Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal.
Aunque la LOPDGDD solo ha derogado expresamente la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (Disposición derogatoria única) y no así el Real Decreto 1720/2007, de 21 de diciembre, mayoritariamente se entiende que sí lo estaría, por cuanto la Disposición derogatoria única de la LOPDGDD establece que "*Asimismo, quedan derogadas cuantas*

contenía el concepto de "*fuentes accesibles al público*"[3] sobre cuyos datos las compañías quedaban exceptuadas de facilitar información a los sujetos afectados por el tratamiento y de obtener su consentimiento. Sin embargo, la excepción quedaba limitada a algunas bases de datos públicas (por ejemplo, censos promocionales, repertorios telefónicos, listado de profesionales - por ejemplo, abogados -, diarios, boletines oficiales y medios de comunicación[4]).

Con la entrada en vigor del RGPD, el concepto de "*fuentes accesibles al público*" quedó abandonado y, aunque algunos operadores del mercado consideran que el concepto "*fuentes accesibles al público*" sigue resultando de aplicación en su cierta medida, la opinión mayoritaria es que no lo hace, principalmente por dos motivos clave:

(a) Por un lado, como hemos adelantado, la disposición derogatoria única de la LOPDGDD rechaza completamente la aplicación de la normativa anterior, así como la normativa que pueda contradecir, oponerse o resultar incompatible con el RGPD.

(b) Por otro lado, el RGPD y la LOPDGDD contienen diversas disposiciones sobre deberes de información y la necesidad de que exista una base jurídica necesaria para el tratamiento de datos personales, así como excepciones y situaciones en las que existen otras bases de legitimación que pueden prevalecer en función de las circunstancias del caso concreto y el uso de datos de fuentes públicas, a diferencia de lo que ocurría en la normativa anterior, no se encuentra entre ellas.

---

disposiciones de igual o inferior rango contradigan, se opongan, o resulten incompatibles con lo dispuesto en el Reglamento (UE) 2016/679 y en la presente ley orgánica".

[3] Artículo 3.(j), 5 y 6 y 28 de la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal, entre otros.

[4] Artículo 7 del Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal.

En este sentido, además, se ha pronunciado la AEPD, que recientemente ha emitido un informe en que da su opinión sobre el uso de datos personales de "fuentes públicas"[5] para analizar la solvencia de una persona. En este informe en que la AEPD se pronuncia sobre la aprobación de un Código de Conducta del sector infomediario, presentado por la Asociación Multisectorial de la Información (ASEDIE) y publicado el pasado abril de 2021, la AEPD entiende que no existe un término legal de "*fuente accesible al público*" y que, por tanto, no existe un interés legítimo automático y general para usar los datos que provengan de estas fuentes. El uso de estas fuentes, por tanto, deberá estar sujeto a los principios de la normativa de protección de datos, incluyendo la necesidad de tener una base jurídica y la obligatoriedad de dar información al usuario sobre el tratamiento de sus datos personales. La base jurídica podrá ser el interés legítimo y, aunque a tal el hecho de que los datos tratados provengan de fuentes de acceso público puede ser tenido en consideración como una circunstancia a la hora de analizar la existencia de un interés legítimo, este hecho no es ni mucho menos determinante de forma automática.

2.3.Obligaciones derivadas de la normativa de protección de datos

Por consiguiente, en tanto que los datos personales a los que se pueda acceder mediante el uso de estas técnicas traten datos personales, se deberá tener en consideración la normativa existente en materia de protección de datos, de igual forma que se aplica para el acceso a cualquier dato personal proveniente de otro tipo de fuentes.

Esto supone, entre otras obligaciones, la necesidad de que exista una **base de legitimación** que habilite este tratamiento, de conformidad con el principio de transparencia y legitimidad del art. 5.1.a) del RGPD. Estas bases legítimas están previstas en el art. 6.1 del RGPD[6], si bien, a efectos de estas técnicas, las únicas

---

[5] Informe de la Agencia Española de Protección de Datos número 0089/2020 https://www.aepd.es/es/documento/2020-0089.pdf

6 En atención a este artículo 6.1. del RGPD, el tratamiento de cualquier dato personal solo será lícito si se cumple al menos una de las siguientes condiciones:

a) "*el interesado dio su consentimiento para el tratamiento de sus datos personales para uno o varios fines específicos;*

bases jurídicas que podrán ser generalmente utilizadas son solo dos: (i) el consentimiento o (ii) el interés legítimo.

Los operadores que utilizan estas técnicas suelen apoyarse masivamente en el interés legítimo, puesto que el consentimiento resulta una labor imposible, si bien, como decimos, la AEPD entiende que su aplicación no puede ser automática y que, si bien en algunos casos podrá utilizarse, en otros no. Por ejemplo, en materia de tecnologías de reconocimiento facial o de análisis de estados anímicos, siempre y cuando no entren dentro de las "*categorías especiales de datos personales*", entendemos que difícilmente su uso podrá apoyarse en este "interés legítimo" y es probable que la AEPD entienda que el tratamiento mediante el uso de estas técnicas es excesivo. En todo caso, para valorar si este interés legítimo puede concurrir, debería realizarse previamente un análisis sobre la ponderación de este interés legítimo.

Si, además, estos datos personales tienen la consideración de "*categorías especiales de datos personales*"[7], nos encontraremos ante un tratamiento de categorías especiales de datos que solo se podrá realizar cuando estemos ante una de las excepciones previstas en el artículo 9.2 del RGPD[8]. En tales casos, su tratamiento

---

b) el tratamiento es necesario para la ejecución de un contrato en el que el interesado es parte o para la aplicación a petición de este de medidas precontractuales;

c) el tratamiento es necesario para el cumplimiento de una obligación legal aplicable al responsable del tratamiento;

d) el tratamiento es necesario para proteger intereses vitales del interesado o de otra persona física;

e) el tratamiento es necesario para el cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos conferidos al responsable del tratamiento;

f) el tratamiento es necesario para la satisfacción de intereses legítimos perseguidos por el responsable del tratamiento o por un tercero, siempre que sobre dichos intereses no prevalezcan los intereses o los derechos y libertades fundamentales del interesado que requieran la protección de datos personales, en particular cuando el interesado sea un niño."

[7] Las categorías especiales de datos personales son aquellos "*datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida o la orientación sexuales de una persona física*".

[8] Las posibles excepciones que legitiman el tratamiento de "categorías especiales de datos" en virtud del artículo 9.2 del RGPD serían únicamente las siguientes:

a) "el interesado dio su consentimiento explícito para el tratamiento de dichos datos personales con uno o más de los fines especificados (excepto en ciertos casos en los que la normativa nacional dispone que el consentimiento explícito no es suficiente, como sucede en España

estaría sujeto a los mecanismos de protección reforzada que tal categoría de datos conlleva, mecanismos que incluyen la necesidad de aplicar una de las excepciones que permiten el tratamiento de estos datos listada en el artículo 9.2 del RGPD, además de la concurrencia de una de las bases jurídicas previstas en el artículo 6.1 del RGPD.

Además, en virtud del principio de transparencia previsto en el art. 5.1.a) del RGPD, también se debe **facilitar información a los sujetos afectados sobre el tratamiento** que se va a realizar y esta información debe facilitarse, al menos, en el plazo máximo de un mes desde que se tuvo acceso a los datos[9].

Si bien existe una excepción general que permite a la empresa no tener que informar al interesado cuando "*la comunicación de dicha información resulte imposible o suponga un esfuerzo desproporcionado*" o en la medida en que el deber de

---

respecto al tratamiento de datos relativos a la ideología, afiliación sindical, religión, orientación sexual, creencias u origen racional o étnico, en aplicación del artículo 9.1 de la LOPDGDD)*;*

b) el tratamiento es necesario para el cumplimiento de obligaciones y el ejercicio de derechos específicos del responsable del tratamiento o del interesado en el ámbito del Derecho laboral y de la seguridad y protección social, en la medida en que así lo autorice el Derecho de la Unión de los Estados miembros o un convenio colectivo (...);

c) el tratamiento es necesario para proteger intereses vitales del interesado o de otra persona física, en el supuesto de que el interesado no esté capacitado, física o jurídicamente, para dar su consentimiento;

d) el tratamiento es efectuado, en el ámbito de sus actividades legítimas y con las debidas garantías, por una fundación, una asociación o cualquier otro organismo sin ánimo de lucro, cuya finalidad sea política, filosófica, religiosa o sindical (...);

e) el tratamiento se refiere a datos personales que el interesado ha hecho manifiestamente públicos;

f) el tratamiento es necesario para la formulación, el ejercicio o la defensa de reclamaciones o cuando los tribunales actúen en ejercicio de su función judicial;

g) el tratamiento es necesario por razones de un interés público esencial, sobre la base del Derecho de la Unión o de los Estados miembros (...);

h) el tratamiento es necesario para fines de medicina preventiva o laboral, evaluación de la capacidad laboral del trabajador, diagnóstico médico, prestación de asistencia o tratamiento de tipo sanitario o social, o gestión de los sistemas y servicios de asistencia sanitaria y social, sobre la base del Derecho de la Unión o de los Estados miembros o en virtud de un contrato con un profesional sanitario (...);

i) el tratamiento es necesario por razones de interés público en el ámbito de la salud pública, como la protección frente a amenazas transfronterizas graves para la salud, o para garantizar elevados niveles de calidad y de seguridad de la asistencia sanitaria y de los medicamentos o productos sanitarios (...);

j) el tratamiento es necesario con fines de archivo en interés público, fines de investigación científica o histórica o fines estadísticos (...)".

[9] Artículo 14.3.a) del RGPD.

información "*pueda imposibilitar u obstaculizar gravemente el logro de los objetivos de tal tratamiento"*[10], esta excepción se aplica de forma restrictiva, por lo que es muy discutible en la mayoría de las situaciones.

Por ejemplo, la Autoridad de Protección de Datos de Polonia impuso una sanción administrativa de 220.000€ en marzo de 2019, a una entidad que llevaba a cabo *web scrapping* sin haber informado a 5.7 millones de sujetos cuyos datos habían sido objeto de tratamiento. La entidad reclamada alegó haber colocado un aviso de privacidad en su sitio web como medida suficiente para notificar a los individuos[11], sin embargo, la autoridad polaca rechazó este argumento y, por tanto, la aplicación de la excepción, al entender que esta medida no era suficiente y los afectados no eran conscientes de que sus datos personales estaban siendo objeto de *web scrapping* y de su tratamiento posterior. Teniendo en cuenta que la AEPD suele mantener una posición restrictiva en la aplicación de este tipo de excepciones, y que en los últimos meses el importe económico de las sanciones por incumplimientos normativos se ha visto incrementado de forma llamativa, es probable que España mantuviera una posición similar a la de Polonia.

Y, aun aplicando esta excepción, se debería adoptar otras medidas apropiadas para proteger los derechos, libertades e intereses legítimos del interesado, incluyendo hacer pública la información (por ejemplo, en la página web).

Internamente, además, es probable que debiese realizarse una **evaluación de impacto en materia de protección de datos** de forma previa a iniciar el tratamiento[12], en que se analicen los riesgos y amenazas a los derechos de los interesados y se establezcan las medidas adicionales que se asumen para aminorar esos riesgos.

---

[10] Art. 14.5.b) del RGPD.
[11] Esta medida aparece recogida incluso junto con la excepción en el art. 14.5.b) del RGPD.
[12] Art. 35 del RGPD y guía de la AEPD de Gestión del riesgo y evaluación de impacto en tratamientos de datos personales, de 29 de junio de 2021.

El cumplimiento de estos deberes parece *cuasi* imposible en la mayoría de las situaciones en las que se desea utilizar este tipo de técnicas, por cuanto se busca precisamente inmediatez y acceso rápido a ingente información, con finalidades muy diversas.

3. Propiedad intelectual: ¿qué pasa con las licencias de uso de los terceros titulares de las páginas web de las cuales se extrae la información?

También se derivan problemas en materia de propiedad intelectual. En esta materia, los principales problemas giran en torno a las licencias de uso de las páginas web que se visitan.

Por un lado, se han dado diversos supuestos a nivel jurisprudencial en los que se han iniciado potenciales acciones civiles por vulneración de derechos sobre los titulares de sitios web -que contienen bases de datos- sobre los que se han utilizado técnicas de *web scrapping* o similares (e.g. *screen scraping*).

A este respecto, deberá tenerse en cuenta si el contenido de la página web que se visita mediante la técnica de *web scrapping* es susceptible de protección vía derechos de propiedad intelectual. Esto solo sucederá cuando se entienda que, por su selección o disposición de contenidos, la base de datos de la página web constituye una "creación intelectual" sobre la que se puedan invocar derechos de autor[13], en cuyo caso, además, la carga de la prueba recaerá sobre el propio titular de la página web que deberá probar que es susceptible de protección y de que la empresa ha utilizado una herramienta de *web scrapping* que le ha permitido copiar el contenido. De nuevo, esto resulta extremadamente difícil en la práctica.

Por otro lado, también ha habido supuestos jurisprudenciales en torno a acciones de incumplimiento del aviso legal/condiciones de uso de la página web. Es frecuente

[13] Entre otros, cabe citar la Sentencia del Tribunal de Justicia de la Unión Europea, de 15 de enero de 2015, y las del Tribunal Supremo de 7 de mayo de 2014 y de 9 de octubre de 2012, en las que se negó esta protección.

que las condiciones de uso de sitios web contengan restricciones a la posibilidad de reproducir, distribuir, ceder, transmitir, modificar o suprimir la información o contenido.

A este respecto, a nivel jurisprudencial, se ha discutido principalmente si: (i) la aceptación de condiciones de uso de un sitio web por la mera navegación es equivalente a una relación contractual como pueda ser un contrato de prestación de servicios; y (ii) en su caso, si al utilizar técnicas de *web scrapping* de forma contraria a dichas condiciones de uso estaríamos ante una infracción de condiciones generales de contratación y, como consecuencia, podrían derivarse responsabilidades[14].

No es un tema claro y, de nuevo, en este supuesto, serían los titulares de los sitios web los que deberían probar el uso de esta técnica sobre los contenidos desplegados en su sitio web y su uso posterior de forma contraria a las condiciones de uso, lo cual, de nuevo, es enormemente difícil en la práctica.

4. Inteligencia artificial y uso de algoritmos: la dificultad añadida de cumplir con los requisitos de transparencia, gobernanza de datos o precisión.

Por último, también existen riesgos desde la óptica regulatoria y ética del uso de algoritmos y técnicas de inteligencia artificial. Actualmente, todos estos sistemas de *web scrapping*, normalmente, utilizan técnicas de inteligencia artificial, como por ejemplo técnicas de aprendizaje automático, que estarían sujetos a la futura normativa de regulación de la inteligencia artificial.

---

[14] Aunque existen dudas al respecto y no existen decisiones de actualidad, cabe citar, de nuevo, la Sentencia del Tribunal Supremo de 7 de mayo de 2014, en la que se determinó que no había infracción contractual: *"No puede confundirse la infracción de un contrato, que exige la previa prestación de consentimiento, con la actuación de quien, con rechazo más o menos expreso de las condiciones impuestas por el titular de una página web y, por vía de hecho, a modo de atajo o pasarela facilita a los consumidores finales el acceso directo a determinados contenidos de la web, sin necesidad de seguir el itinerario de navegación diseñado por su titular, durante el cual se ofertan al usuario pluralidad de productos o servicios más o menos relacionados con el vuelo que el consumidor final pretende contratar y que no podrán ser aceptadas al ser desconocidas."* De forma previa, se pronunció en términos similares la Audiencia Provincial de Barcelona, en su Sentencia de 17 de diciembre de 2009.

Así, el pasado 21 de abril de 2021, la Comisión Europea publicó su propuesta de marco jurídico regulatorio sobre inteligencia artificial ("**Propuesta de regulación de IA**")[15], sobre el que ya han emitido algunas enmiendas y que propugna los principios de transparencia, gobernanza de datos, trazabilidad y exigencias de precisión y seguridad, que ya venían proclamándose desde el 2018 y 2019, con la publicación de la Estrategia Europea de Inteligencia Artificial[16], la Comunicación sobre la construcción de confianza en una inteligencia artificial centrada en el ser humano[17] y las directrices para conseguir una inteligencia artificial "*confiable*"[18].

Esta Propuesta de regulación de IA establece obligaciones de transparencia, trazabilidad y gobernanza de datos, seguridad y precisión para aquellos sistemas de inteligencia artificial que entren dentro de la consideración de "alto riesgo"[19]. Aunque no resulta aún claro cómo se interpretará este deber de "transparencia" y qué información deberá entregarse a los usuarios, deberá incluir en todo caso información sobre la finalidad, nivel de precisión, solidez y ciberseguridad, su funcionamiento, especificaciones relativas a los datos de entrada, datos de entrenamiento, validación y prueba usados, entre otros. Asimismo, la AEPD ya se ha pronunciado al respecto y ha indicado que esta información deberá incluir información significativa sobre la existencia, lógica aplicada, importancia y consecuencias, incluyendo información sobre el detalle de los datos empleados,

---

[15] 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 (https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206)

[16] AI Strategy COM(2018) 237

[17] Communication on Building Trust in human-Centric AI COM(2019) 168

[18] Guidelines on trustworthy AI, publicadas por el grupo de expertos en IA establecidos por la Comisión Europea.

[19] Esto es, cuando, de conformidad con el provisional art. 6 de la Propuesta de regulación de IA el sistema de IA esté destinado a ser utilizado como componente de seguridad de un producto contemplado en la legislación armonizada de la UE o sea en sí mismo uno de esos productos y el producto deba someterse a una evaluación de la conformidad realizada por un organismo independiente para su introducción en el mercado o puesta en servicio; o cuando estemos ante uno de los sistemas listados en la propia Propuesta de regulación de IA.

importancia relativa de cada uno de ellos, calidad de los datos de entrenamiento y tipo de patrones, perfilados, valores de precisión o error, etc.[20].

Pero es que, además, y sin perjuicio de que la mencionada Propuesta de regulación de IA no haya sido aún aprobada, los algoritmos están siendo objeto de distintos pronunciamientos a nivel europeo en los que se aboga por su transparencia y se sanciona su uso de forma "*discriminatoria*" o sin que se haya ofrecido la suficiente información a los usuarios[21].

Estos principios también aparecen en las últimas declaraciones de derechos digitales que hemos ido comentando.

La necesidad de que este tipo de técnicas sean "transparentes" genera problemas cuando nos damos cuenta, por ejemplo, de que la mayoría de estas técnicas usan fuentes públicas, librerías, códigos abiertos, etc., que dificultan en gran medida poder cumplir con este requisito de "transparencia", puesto que ni los propios programadores saben en muchas ocasiones cuál es la lógica o cómo toma decisiones el algoritmo.

5. Gobernanza europea de datos: ¿pero no busca precisamente ese uso de la información de forma sencilla la propuesta de reglamento relativo a la gobernanza europea de datos, o Ley de Gobernanza de Datos[22]?

---

[20] Guía de la AEPD sobre la Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial, publicado en febrero de 2020 (https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf); junto con la Guía de la AEPD sobre los Requisitos para Auditorías de Tratamientos que incluyan IA, publicado en enero de 2021.

[21] Sentencia del *Tribunale Ordinario di Bologna* de 31 de diciembre de 2020 contra Deliveroo, por el uso del algoritmo "Frank" que clasificaba a los *riders* en función de ciertos parámetros y penalizaba por igual a los empleados que se hubieran ausentado del trabajo por razones injustificadas de aquellos que lo habían hecho por razones justificadas. También ha habido otro pronunciamiento del tribunal de Casación de Italia, de 25 de mayo de 2021 (*Garante per la Protezione dei Dati Personali v. Associazione Mevaluate Onlu*), en torno a la necesidad de que los algoritmos sean transparentes para que el consentimiento de un usuario sea libre. De igual forma, la autoridad de protección de datos coreana ha impuesto una multa a Scatter Lab por el tratamiento indiscriminado de datos por una tecnología chatbot llamada Iruda.

[22] Propuesta de Reglamento del Parlamento Europeo y del Consejo relativo a la gobernanza europea de datos (Ley de Gobernanza de Datos), COM(2020) 767 final, publicado el 25 de noviembre de 2020.

Actualmente, tenemos sobre la mesa una propuesta de Reglamento del Parlamento Europeo y del Consejo relativo a la gobernanza europea de datos ("**Ley de Gobernanza de Datos**"), cuyo objetivo es precisamente facilitar el acceso de las pequeñas y medianas empresas a los datos – personales o no -, para el desarrollo de soluciones innovadoras y para beneficiar a la sociedad de los grandes volúmenes de datos recogidos por el sector público con cargo a los presupuestos generales, mediante la disponibilidad de los datos para su reutilización, el aumento de la confianza en los intermediarios de datos y el refuerzo de los mecanismos de intercambios.

Esta propuesta de Ley de Gobernanza de Datos está acompañada de la actual Directiva de datos abiertos[23], cuya aplicación solo está enfocada a aquellos datos no sujetos a ningún tipo de derechos.

A priori, los objetivos de esta normativa parecen apoyar el uso de este tipo de técnicas para el acceso a información para su reutilización y, a la vez, parecen también contraponerse a gran parte de los principios existentes en otras materias ya mencionadas, como la protección de datos, la propiedad intelectual o incluso los secretos empresariales.

La relación entre esta normativa y las demás regulaciones no es clara y deberemos esperar a su aprobación, para profundizar en mayor medida en su interacción y en cómo podemos articular, por ejemplo, los principios de transparencia y legitimación versus la disponibilidad y el intercambio abierto de datos. Por el momento, lo que resulta claro es la existencia de una fricción patente que puede conllevar grandes discusiones en la práctica.

6. Conclusiones

[23] DIRECTIVA (UE) 2019/1024 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 20 de junio de 2019 relativa a los datos abiertos y la reutilización de la información del sector público.

El uso de estas técnicas de *web scrapping* e inteligencia artificial que usan información de fuentes públicas presenta numerosos problemas y fricciones en diversas materias, sobre todo en relación con el cumplimiento de los principios de transparencia y autonomía.

Problemas estos que se suman a la extremada dificultad de demostrar nada, lo que conlleva que la mayoría de las empresas que presten estos servicios, incluso aquellas ubicadas en la Unión Europea, tiendan a “ignorar” u “obviar” los inconvenientes legales derivados de su uso. Y así, el uso de estas técnicas es, cada vez, más significativo.