

Transparencia e inteligencia artificial en Europa. Una relación normativa incierta

Susana Beltrán García

Universitat Autònoma de Barcelona

Susana.Beltran@uab.cat

ORCID: 0000-0002-3846-2214



Recepción: 31/1/2025

Aceptación: 05/6/2025

Publicación: 26/6/2025

Cita recomendada: BELTRÁN GARCÍA, S. (2025). "Transparencia e Inteligencia Artificial en Europa. Una relación normativa incierta". *Quaderns IEE: Revista de l'Institut d'Estudis Europeus*, 4(2), 27-58.

DOI: <<https://doi.org/10.5565/rev/quadernssee.120>>

Resumen

En los sistemas de inteligencia artificial (IA), la exigencia de transparencia es fundamental para que los usuarios comprendan que están actuando con este tipo de tecnología y conozcan sus riesgos. Sin embargo, el Reglamento Europeo de Inteligencia Artificial no define explícitamente qué es transparencia, lo que ya ocasiona incertidumbre sobre su contenido, alcance e implementación. En efecto, a pesar de los esfuerzos del legislador europeo por regular la IA de manera que se preserven los derechos de las personas, piedra angular en toda esta materia, persisten dudas sobre cómo aplicar la transparencia de forma que contribuya efectivamente a dicha protección. En este artículo se abordan los problemas de carácter normativo y técnico que surgen por la falta de concreción del concepto de transparencia en los sistemas de IA, así como la necesidad de clarificar su aplicación. En este sentido, la Unión Europea tiene ante sí una oportunidad para delimitar los contornos jurídicos de la transparencia y contribuir, a través de sus reglas, a la gobernanza mundial de la IA.

Palabras clave: Transparencia; Inteligencia artificial; Rendición de cuentas; Reglamento Europeo de Inteligencia Artificial; Trazabilidad; Explicabilidad.

Resum. *Transparència i intel·ligència artificial a Europa. Una relació normativa incerta*

En els sistemes d'intel·ligència artificial (IA), l'exigència de transparència és fonamental perquè els usuaris comprenguin que estan interactuant amb aquest tipus de tecnologia i en coneguin els riscos. Tanmateix, el Reglament Europeu d'Intel·ligència Artificial no defineix explícitament què s'entén per transparència, cosa que ja genera incertesa sobre el seu contingut, abast i implementació. En efecte, malgrat els esforços del legislador europeu per regular la IA de manera que es preservin els drets de les persones —pedra angular en tota aquesta matèria—, persisteixen dubtes sobre com aplicar la transparència de manera que contribueixi efectivament a aquesta protecció. En aquest article s'aborden els problemes de caràcter normatiu i tècnic que sorgeixen per la manca de concreció del concepte de transparència en els sistemes d'IA, així com la necessitat de clarificar-ne l'aplicació. En aquest sentit, la Unió Europea té al davant una oportunitat per delimitar els contorns jurídics de la transparència i contribuir, a través de les seves normes, a la governança mundial de la IA.

Paraules clau: Transparència; Intel·ligència artificial; Rendició de comptes; Reglament Europeu d'Intel·ligència Artificial; Traçabilitat; Explicabilitat.

Abstract. *Transparency and artificial intelligence in Europe. An uncertain regulatory relationship*

In artificial intelligence (AI) systems, the demand for transparency is essential for users to understand that they are interacting with this type of technology and to be aware of its risks. However, the Artificial Intelligence Act does not explicitly define what transparency means, which already creates uncertainty regarding its content, scope, and implementation. Indeed, despite the efforts of the European legislator to regulate AI in a way that safeguards individuals' rights — a cornerstone of this entire field — doubts remain about how to apply transparency in a manner that effectively contributes to such protection. This article addresses the normative and technical challenges that arise from the lack of clarity surrounding the concept of transparency in AI systems, as well as the need to clarify its application. In this regard, the European Union faces an opportunity to define the legal contours of transparency and to contribute, through its rules, to the global governance of AI.

Keywords: Transparency; Artificial intelligence; Accountability; Artificial Intelligence Act; Traceability; Explainability.

Sumario

1. Introducción
 2. Concepto de transparencia e inteligencia artificial en la legislación europea e internacional
 3. Obligaciones de transparencia según el Reglamento de Inteligencia Artificial
 4. Transparencia e incertezas
 5. Conclusiones: Riesgos por exceso o defecto de transparencia
 6. Referencias
-

1. INTRODUCCIÓN

La inteligencia artificial (IA) está cada vez más presente en nuestra vida cotidiana, muchas veces incluso sin que seamos plenamente conscientes de ello. Aunque el número de personas que saben manejar estas herramientas y aplicarlas a usos concretos va en aumento —como ocurre con *ChatGPT*, de uso generalizado, o plataformas especializadas como *Sof-IA* o *ScreenApp.AI*, diseñadas para el ámbito jurídico—, a menudo se pasan por alto los riesgos asociados a su utilización, así como la veracidad de los resultados que generan.

En muchos casos, el razonamiento dominante es simple: si la herramienta funciona y se sabe utilizar, eso parece ser suficiente. Sin embargo, esta confianza acrítica puede encubrir errores, sesgos o consecuencias no deseadas, especialmente cuando se delegan decisiones relevantes a sistemas cuya lógica interna no se comprende del todo.¹ Así, esto ocurre en la denominada “caja negra” (*black box*) de la IA, que es el proceso interno que media entre la entrada al sistema, en donde se realiza el volcado y entrenamiento del modelo con datos, y la salida, en el que no se sabe qué ha pasado dentro y por qué se obtienen tales resultados. En efecto, en el volcado pueden haberse utilizado datos personales sin que seamos conscientes y sin saber qué uso se va a dar de los mismos y en cuanto al resultado se puede dar por fiable una información que quizás no lo sea.

Por ello es importante entender qué pasa en la caja negra y uno de los principios esenciales a tal fin es el de la transparencia. No obstante, abordar en estos momentos la transparencia y su relación con la IA conlleva ciertas dificultades. En primer lugar, los conceptos, qué son los sistemas de IA, a quiénes van dirigidas las obligaciones de transparencia y quiénes pueden exigir las. O cuando la propia legislación europea prevé complementarse con la normativa nacional e internacional ¿se está hablando de lo

¹ Entre otros; https://www.eca.europa.eu/ECAPublications/SR-2024-08/SR-2024-08_EN.pdf; https://www.epdata.es/datos/inteligencia-artificial-espana-union-europea-empresas/727?utm_source=chatgpt.com; <https://www.itreseller.es/en-cifras/2024/06/dos-de-cada-tres-espanoles-ya-estan-familiarizados-con-la-inteligencia-artificial>

mismo? En segundo lugar, la comprensión sobre el funcionamiento de los sistemas de IA: no es posible justificar por qué es necesaria la transparencia, si no se entiende qué es y cómo se desarrollan los sistemas de IA. Estas dificultades se extienden a los ingenieros que diseñan dichos modelos y que no conocen los riesgos que entrañan su puesta en el mercado para la protección de los derechos de las personas, como los datos personales, la seguridad o el medio ambiente. Para los juristas, que sí que entienden dichos riesgos, pero que escapan a su comprensión múltiples facetas de cómo funcionan dichos sistemas.² Para los usuarios que interactúan con la IA, qué es lo que tienen que saber acerca de ella y por qué deben exigir transparencia. Para las autoridades nacionales y europeas que deben velar por el cumplimiento de la transparencia, si no entienden su necesidad. En tercer lugar, la imposibilidad de prever todas las consecuencias derivadas de la aplicación de los sistemas de IA. Si algo se ha evidenciado, especialmente con la denominada IA generativa, es que no siempre se pueden predecir sus resultados: si seguirá una secuencia lógica o cambiará sus códigos, en uso de su autonomía, una vez ya se ha puesto en marcha. Finalmente, y debido a las razones expuestas, puede ser que la legislación que se toma de base, en particular el Reglamento de inteligencia artificial,³ quede en desuso nada más nacer.

Con todo, el vertiginoso ritmo de evolución de la inteligencia artificial, junto con sus posibles consecuencias, obliga a abordar diversos aspectos relacionados con ella —como la transparencia— sin disponer aún de todas las claves ni de una visión completa de sus implicaciones. Esta realidad, de hecho, continúa siendo objeto de numerosas críticas. Sirva de ejemplo, las declaraciones del entonces ministro de finanzas francés, Bruno Le Maire, que en una declaración sobre la inteligencia artificial, en febrero de 2024, advirtió de los riesgos que corre la Unión Europea al pretender regular la IA en un ámbito que precisa avanzar y no poner barreras legales sobre temas que todavía se desconocen: “Réguler, c'est très bien, mais en règle générale, quand on commence par réguler une innovation, c'est qu'on ne la maîtrise pas. Donc je préfère que nous innovions pour ensuite encadrer, réguler et mettre toutes les barrières qui sont nécessaires”.

La realidad es que la legislación europea sobre IA se ha materializado a pesar de las incertidumbres que planean sobre ella y que empiezan por el propio concepto de IA. En este artículo, para centrar el tema, se va a partir de la definición de la IA dispuesta en el Reglamento de inteligencia artificial:

² En este sentido, publicaciones que clarifiquen contenidos dirigidas a todo tipo de públicos son esenciales para no perderse en el entorno digital, y en particular, con los sistemas de IA. Entre otros y por su actualidad, véase Blasi, Cristina y Cañabate, Josep; *Legislación y derecho digital para no juristas*, UAB Servei de Publicacions, 2024

³ Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial) (Texto pertinente a efectos del EEE)

Sistema de IA»: un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales. (art. 3.1)

2. CONCEPTO DE TRANSPARENCIA E INTELIGENCIA ARTIFICIAL EN LA LEGISLACIÓN EUROPEA E INTERNACIONAL

2.1. La Declaración europea sobre derechos y principios digitales de 2022

Un antecedente claro del Reglamento de inteligencia artificial, en el que se aborda la cuestión de la transparencia, es la Declaración Europea sobre los Derechos Digitales de 2022. Este documento establece de forma contundente que la inteligencia artificial debe ser una herramienta al servicio de las personas, cuyo objetivo último sea el de promover y aumentar el bienestar humano. En este sentido, la transparencia se vincula estrechamente con la libertad de elección en el uso de la inteligencia artificial. En concreto, se destaca el compromiso de la Unión Europea por garantizar un nivel adecuado de transparencia en relación con el uso de algoritmos y sistemas de IA, de modo que las personas puedan utilizarlos con conocimiento de causa y estén debidamente informadas cuando interactúan con ellos.⁴ Todas las personas deben poder beneficiarse de las ventajas que ofrecen los sistemas algorítmicos y la inteligencia artificial, incluyendo la capacidad de tomar decisiones autónomas en el entorno digital. Al mismo tiempo, deben estar protegidas frente a los posibles riesgos y perjuicios que puedan afectar a su salud, su seguridad o sus derechos fundamentales.

Es decir, la exigencia de transparencia se concibe como un mecanismo clave para proteger los derechos de las personas frente a los posibles riesgos que los sistemas de inteligencia artificial puedan generar. Al mismo tiempo, busca favorecer una experiencia más consciente y satisfactoria en la interacción con estos sistemas. No obstante, como veremos más adelante, lo que *a priori* parece una premisa tan evidente e imprescindible, en la práctica está plagado de dificultades y desafíos.

La Declaración menciona la transparencia, pero con otras vinculaciones más generales, no específicas de la IA. En particular, la considera una forma de solidaridad y de inclusión para lograr condiciones de trabajo justas e igualmente, la contempla para asegurar entornos digitales justos. En concreto, respecto a la libertad de elección de servicios *online*. Finalmente, demanda promover la transparencia para conseguir una mayor confianza en la tecnología.⁵

⁴ Punto 9, Capítulo III *Libertad de elección*, de la Declaración Europea sobre Derechos y Principios digitales, diciembre 2022

⁵ Posteriormente, el Reglamento de inteligencia artificial sí las recoge específicamente como la vinculación de la transparencia para mejorar las condiciones laborales en el trabajo en plataformas digitales (Considerando 9)

La Declaración se enmarca en la denominada “Década Digital” promovida por la Unión Europea. Así se realizan informes periódicos para analizar los progresos alcanzados y dar seguimiento de la aplicación de esta. En concreto, el Informe sobre el Estado de la Década Digital de 2024⁶ contempla la adopción del Reglamento de inteligencia artificial como uno de sus logros y remite a su articulado para el desarrollo de lo que significa transparencia e IA.⁷ No obstante, insiste en la idea de que la transparencia implica la obligación de informar a las personas cuando están interactuando con un sistema de inteligencia artificial.

En efecto, ¿cuál es el valor jurídico de la Declaración Europea sobre Derechos y Principios Digitales de 2022? Se trata de una declaración conjunta, de carácter político, suscrita por el Parlamento Europeo, el Consejo y la Comisión, mediante la cual estas instituciones se comprometen a tenerla en cuenta en el desarrollo de sus acciones. La propia declaración señala que su promoción e implementación constituyen un compromiso político, cuya responsabilidad recae tanto en la Unión Europea como en los Estados miembros, dentro del ámbito de sus respectivas competencias.

Al tratarse de un instrumento de naturaleza política y no jurídica, su nivel de cumplimiento es más flexible y su seguimiento, más impreciso, en comparación con lo que implicaría su inclusión en una norma legal vinculante. Por este motivo, el Informe sobre la Década Digital de 2024 destaca la adopción del Reglamento de inteligencia artificial como una concreción normativa de los objetivos establecidos en dicha declaración.

No obstante, para comprender qué se entiende por transparencia en los sistemas de inteligencia artificial desde una perspectiva normativa, no basta con atender al Reglamento de inteligencia artificial, que entró en vigor en agosto de 2024. También debe considerarse la *Convención Marco sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho*, adoptada en el seno del Consejo de Europa el 5 de septiembre de 2024. Aun así, como veremos, sigue siendo difícil delimitar con precisión el concepto de transparencia y sus alcances.

2.2. El Reglamento de inteligencia artificial

El Reglamento no ofrece una definición explícita de lo que se entiende por transparencia. El artículo 3, dedicado a las definiciones —y que incluye un amplio número de conceptos— omite cualquier referencia directa a la transparencia.⁸ Por toda explicación, en el Considerando 27 del Reglamento se menciona el Informe del grupo de expertos de alto nivel sobre IA de 2019, que estableció siete principios éticos no

⁶ COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES Estado de la Década Digital 2024 COM/2024/260 final

⁷ Anexo IV del Paquete sobre el Estado de la Década Digital de 2024
<https://ec.europa.eu/newsroom/dae/redirection/document/106722>

⁸ En total son 68 definiciones contenidas en el artículo 3.

vinculantes para una IA fiable, incluida la transparencia.⁹ En concreto se entiende por transparencia:

los sistemas de IA se desarrollan y utilizan de un modo que permita una trazabilidad y explicabilidad adecuadas, y que, al mismo tiempo, haga que las personas sean conscientes de que se comunican o interactúan con un sistema de IA e informe debidamente a los responsables del despliegue acerca de las capacidades y limitaciones de dicho sistema de IA y a las personas afectadas acerca de sus derechos” (Considerando 27).¹⁰

Según este informe, la transparencia en los sistemas de inteligencia artificial abarca diversos elementos clave, como los datos utilizados, el propio sistema, y los modelos de negocio que lo sustentan. Además, debe reunir una serie de características específicas que garanticen su eficacia y utilidad:

Trazabilidad. La transparencia implica tener en cuenta y comprender los datos y procesos que conducen a que un sistema de inteligencia artificial adopte una determinada decisión. Estos elementos incluyen la recopilación y el etiquetado de los datos, así como los algoritmos empleados en el desarrollo del sistema. El objetivo es que todos estos componentes queden debidamente documentados, de modo que sea posible ejercer la trazabilidad y, en consecuencia, identificar qué ha fallado cuando se produce una decisión errónea por parte del sistema. Este mecanismo no solo permite analizar errores pasados, sino que también contribuye a prevenirlos en el futuro. En este sentido, la trazabilidad se convierte en una herramienta esencial para garantizar tanto la auditabilidad como la explicabilidad de los sistemas de IA.

Explicabilidad. Significa la capacidad de poder explicar tanto los procesos técnicos de un sistema de IA como las decisiones humanas que están asociadas. La gran dificultad reside en que las decisiones que adopte un sistema de IA han de ser comprensibles para las personas y además han de poder ser rastreadas. El informe hace hincapié en que hay que buscar un equilibrio entre mejorar la explicabilidad de un sistema (lo que puede reducir su precisión) o una mayor precisión de este (a costa de la explicabilidad). Según los expertos, cuando un sistema de IA pueda impactar de forma significativa en la vida de las personas, debería ser posible exigir una explicación del proceso de toma de decisiones del sistema de IA y que sea inteligible. Es decir, tal explicación tendría que adaptarse al nivel de especialización de la persona afectada, que puede ser, o no, experta en la materia, o ser un regulador o un investigador. Asimismo, habría que disponer de explicaciones sobre la manera en que el sistema de IA condiciona el proceso de toma de decisiones de la organización, sobre las decisiones de diseño del sistema y sobre la lógica subyacente a su despliegue. Esto que supone una

⁹ Estos siete principios éticos para la IA son: 1. Acción y supervisión humanas 2. Solidez técnica y seguridad. 3. Gestión de la privacidad y de los datos. 4. Transparencia. 5. Diversidad, no discriminación y equidad. 6. Bienestar social y ambiental. 7. Rendición de cuentas Grupo de expertos de alto nivel sobre inteligencia artificial; *Directrices éticas para una IA fiable*, abril de 2019, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

¹⁰ Como se ha señalado, el Considerando 27 del Reglamento recoge lo que es transparencia según el informe de expertos, es decir, no es una definición propia.

garantía de transparencia del modelo de negocio, tal como observa el grupo de expertos en su informe, puede ser un obstáculo para quienes quieren emprender en este ámbito (Comisión Europea, 2019).

Comunicación. Las personas tienen que saber que están interactuando con un sistema de IA. Es más, en el informe se señala que cuando sea pertinente, se debería ofrecer al usuario la posibilidad de decidir si quiere interactuar con un sistema de IA o con otra persona. De este modo se garantizaría el respeto de los derechos fundamentales. Por otra parte, habría que informar, según el grupo de expertos, a los profesionales o a los usuarios finales de las capacidades del sistema de IA. Esta información debería incluir aclaraciones sobre su nivel de precisión, así como de sus limitaciones. En el Reglamento esta obligación de comunicación va dirigida a las personas que interactúan con un sistema de IA y a los responsables del despliegue (Comisión Europea, 2019).

Como puede observarse, la transparencia y las características que se derivan de ella, tal como se mencionan en el Reglamento, no tienen carácter vinculante, ya que remiten a un informe que carece de fuerza jurídica. De hecho, la propia norma establece que la transparencia debe ser tenida en cuenta en la elaboración de códigos de conducta, los cuales, a su vez, tampoco son de cumplimiento obligatorio:

(referido también a la transparencia) La aplicación de esos principios debe traducirse, cuando sea posible, en el diseño y el uso de modelos de IA. En cualquier caso, deben servir de base para la elaboración de códigos de conducta en virtud del presente Reglamento. Se anima a todas las partes interesadas, incluidos la industria, el mundo académico, la sociedad civil y las organizaciones de normalización, a que tengan en cuenta, según proceda, los principios éticos para el desarrollo de normas y mejores prácticas voluntarias (Considerando 27).

Ahora bien, sí es posible establecer la obligación de transparencia en función del nivel de riesgo que presente el sistema de inteligencia artificial. En este sentido, el artículo 50 del Reglamento establece obligaciones específicas de transparencia tanto para los proveedores como para los responsables de su despliegue. En particular, la figura del proveedor se encuentra definida en el propio Reglamento, delimitando con precisión sus responsabilidades en relación con la transparencia del sistema:

una persona física o jurídica, autoridad pública, órgano u organismo que desarrolle un sistema de IA o un modelo de IA de uso general o para el que se desarrolle un sistema de IA o un modelo de IA de uso general y lo introduzca en el mercado o ponga en servicio el sistema de IA con su propio nombre o marca, previo pago o gratuitamente (art. 3.3).

Por su parte, el responsable del despliegue:

una persona física o jurídica, o autoridad pública, órgano u organismo que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional (art. 3.4).

Como se verá en las páginas siguientes, por una parte, el Reglamento sí establece obligaciones de transparencia para determinadas figuras, como los proveedores y responsables del despliegue de sistemas de IA. Por otra, se contempla que las personas físicas que interactúan con estos sistemas sean las destinatarias de dicha transparencia y, en su caso, puedan ejercer los derechos que les correspondan.

Además, el Reglamento establece normas armonizadas en materia de transparencia aplicables a determinados sistemas de inteligencia artificial (art. 1.d), y remite, a su vez, a otras normativas europeas relevantes, como el Reglamento (UE) 2016/679 sobre protección de datos. En este marco, las prácticas de gestión y gobernanza de datos personales deben garantizar la transparencia respecto al propósito original de la recopilación de dichos datos (Considerando 67). De hecho, este reglamento considera la transparencia como uno de los principios fundamentales aplicables al tratamiento de datos personales (art. 5.1).

A partir de lo analizado hasta el momento, y a pesar de que el Reglamento no proporciona una definición explícita de transparencia, este concepto aparece mencionado en varios de sus artículos, o bien se remite a otras normativas europeas donde la transparencia se reconoce como un principio fundamental —como ocurre con el Reglamento de Protección de Datos—, o incluso a documentos no vinculantes, como las *Directrices Éticas para una IA Fiable*. Esta dispersión de referencias genera una notable incertidumbre respecto al valor jurídico, al contenido concreto del principio de transparencia y a su articulación con otras normas internacionales aplicables.

2.3. La Convención marco sobre inteligencia artificial, derechos humanos, democracia y Estado de derecho

La *Convención marco sobre inteligencia artificial, derechos humanos, democracia y Estado de derecho*, elaborada en el ámbito del Consejo de Europa, aunque todavía no ha entrado en vigor, prevé que la Unión Europea sea algún día una de las partes a la espera de que proceda a su ratificación.¹¹

En particular, este tratado internacional se centra en asegurar la protección en los sistemas de IA, de los derechos humanos, la democracia y el Estado de derecho, principios que son, asimismo, valores de la Unión Europea (art. 2 del Tratado de la UE). A diferencia del Reglamento, el enfoque no se centra en la prevención del riesgo, sino en la salvaguarda de determinados valores fundamentales. El articulado parte de principios básicos que, si bien aún están poco desarrollados, no debe olvidarse que se trata de una Convención Marco de carácter general. En este sentido, constituye un punto de partida normativo relevante para la regulación de la inteligencia artificial en el ámbito internacional. Su potencial de aplicación es amplio, ya que permite la ratificación por parte de Estados no miembros del Consejo de Europa. De hecho, entre

¹¹ La Unión Europea firmó la citada Convención marco, previa a su ratificación, el 5 de septiembre de 2024

los países signatarios se encuentran Estados Unidos, Israel, Reino Unido, Japón, México, entre otros.

Así, dentro de este marco normativo, la transparencia también ocupa un lugar relevante. Se contempla en el Capítulo III como uno de los principios relativos a las actividades dentro del ciclo de vida de los sistemas de IA y va ligada al principio de vigilancia. En concreto, el tratado dispone que las partes adoptarán o tomarán medidas para asegurar un nivel adecuado de transparencia y vigilancia teniendo en cuenta el contexto y los riesgos respecto de las actividades que se desarrollen a lo largo del ciclo de vida de un sistema de IA. Se incluye también como un requisito más de cumplimiento de la transparencia y vigilancia: la identificación de que los contenidos han sido generados por un sistema de IA (art. 8).

Nótese que, a diferencia del Reglamento de inteligencia artificial, la transparencia según la Convención marco adquiere la categoría de principio. Ahora bien, según su Informe Explicativo, que interpreta la Convención marco, dichos principios —no sólo el de transparencia— tienen un carácter generalista a fin de que puedan implementarse con cierta flexibilidad atendiendo al derecho interno, de acuerdo con los compromisos internacionales adquiridos por cada una de las partes, y a los cambios rápidos que en esta materia puedan suscitarse. Se trata, entonces, de principios orientadores que informan de los elementos que hay que tener en cuenta en la IA.

Respecto de la transparencia, se prevé que los procesos de toma de decisiones y el funcionamiento general de los sistemas de IA han de ser comprensibles y accesibles para los actores de IA pertinentes y para las partes interesadas. Ello supone que, en algunos casos, haya que proporcionar información adicional, como la de los algoritmos utilizados, sin perjuicio de que se tengan en cuenta cuestiones de seguridad, propiedad comercial e intelectual, entre otros. Además, la transparencia puede implicar, según corresponda, informar a las personas interesadas o al público en general sobre los detalles de los datos utilizados para crear, capacitar y operar el sistema, así como información de cómo se protegen los datos personales, el propósito del sistema y el modo en que fue diseñado, probado y desplegado. La transparencia también debería abarcar la comunicación a los actores interesados sobre el procesamiento de la información, los tipos y niveles de automatización utilizados para tomar decisiones, así como los riesgos asociados con el uso del sistema de IA. Ello facilitaría que las partes con intereses legítimos —incluidos los titulares de derechos de autor— ejerzan y hagan cumplir sus derechos de propiedad intelectual. El Informe explicativo es claro respecto a la cuestión de la "transparencia algorítmica" a la que define como "la apertura sobre el propósito, la estructura y las acciones subyacentes de un sistema impulsado por algoritmos".

Por otro lado, y aunque no se mencione explícitamente, las exigencias de los artículos 14, *recursos*, y 15, *salvaguardas procesales*, están conectadas con el principio de transparencia. De hecho, en su Informe Explicativo así lo señala. Sorprende, no obstante, el detalle con el que se desarrolla el contenido del tratado y lo escueto que es

su articulado. En lo concerniente al artículo 14, la Convención marco dispone que las Partes adoptarán las medidas necesarias para garantizar que la información importante sobre los sistemas de IA que puedan afectar a los derechos humanos y su uso esté documentada, permita que los organismos autorizados puedan acceder a dicha información, y si procede, comunicarlo a las personas afectadas. Por su parte, el artículo 15 conmina, entre otras salvaguardas, a que las partes notifiquen a las personas, cuando sea pertinente, que están interactuando con un sistema de IA. En el Informe Explicativo, dichas indicaciones que son de obligado cumplimiento para las partes suponen que han de adoptar las medidas necesarias para que las personas afectadas puedan interponer recursos bien documentados, es decir, han de poder acceder a la información de por qué el sistema de IA ha incidido en sus derechos. Por tanto, deben ser recursos accesibles y eficaces para que las personas afectadas puedan ejercer sus derechos frente a posibles vulneraciones. La Convención marco prevé excepciones a las obligaciones de transparencia por razones de orden público, seguridad y otros intereses importantes en la misma línea de las excepciones previstas en tratados internacionales adoptados en el Consejo de Europa sobre derechos humanos.

Interesa destacar que, en el Informe Explicativo, se presentan como características esenciales de la transparencia: la explicabilidad y la interpretabilidad. La primera se refiere la capacidad de dar explicaciones comprensibles sobre por qué un sistema de IA ofrece información, contenido, recomienda o decide sobre cuestiones sensibles como la salud, las finanzas o la inmigración. Se trata de entender el razonamiento que hay detrás de las decisiones adoptada por un sistema de IA. La segunda, consiste en la capacidad de entender cómo un sistema de IA efectúa sus predicciones o toma decisiones. Es decir, de qué forma los resultados de los sistemas de IA pueden hacerse comprensibles tanto para expertos como para no expertos. Ello afecta al proceso, a entender cómo funciona internamente, la lógica y los procesos de adopción de decisiones de los sistemas de IA.

El Informe Explicativo prosigue que el cumplimiento del principio de transparencia dependerá del sistema de IA utilizado, el contexto de su uso o función, los actores o las partes interesadas afectadas. Tales medidas pueden incluir tener que registrar la procedencia de los datos, las metodologías de capacitación, la validez de las fuentes de datos, la documentación y la transparencia sobre la capacitación, los datos de prueba y validación utilizados, los esfuerzos de mitigación de riesgos y los procesos y decisiones implementados. El objetivo final es que se entienda cómo se derivan los resultados del sistema de IA y cómo impactan en los derechos humanos, la democracia y el estado de derecho. En cualquier caso, una mayor o menor exigencia de transparencia va ligada al riesgo que los sistemas de IA pueden representar para los valores preservados por el tratado.

En este contexto, cabe preguntarse qué norma prevalecerá una vez que los Estados miembros y la propia Unión Europea hayan ratificado la Convención Marco en relación con el Reglamento de inteligencia artificial. *A priori*, no debería existir conflicto

normativo, ya que corresponde a los Estados miembros garantizar el cumplimiento de la Convención, y se parte del supuesto de que no adoptarán ni aplicarán medidas que resulten incompatibles con el Reglamento europeo. Por su parte, cuando la Unión Europea pase a formar parte de este tratado, lo que previsiblemente ocurrirá en el futuro, se reforzará aún más la coherencia normativa y se eliminarán eventuales incompatibilidades entre ambos instrumentos. En la misma dirección, y no menos importante, la Convención marco prevé que las partes que sean miembros de la Unión Europea aplicarán, en sus relaciones mutuas, las normas de la Unión Europea que regulen las materias comprendidas en el ámbito de la Convención marco sin perjuicio del objeto y fin del presente Convenio y sin perjuicio de su plena aplicación con otras partes. Lo mismo se aplica a las otras partes en la medida en que estén sujetas a dichas normas (art. 27). Como ya se ha mencionado, la parquedad del articulado y el carácter generalista de las obligaciones relativas a la transparencia y otros principios permiten una mayor flexibilidad normativa, lo que facilita su armonización con otras regulaciones y contribuye a evitar posibles contradicciones legislativas.

Con todo, no puede descartarse que, una vez ambas normativas hayan entrado en vigor y avance la implementación e innovación de los sistemas de inteligencia artificial, surjan zonas grises desde el punto de vista jurídico. En este sentido, la cláusula de cooperación internacional prevista en la Convención Marco puede contribuir a mitigar la aparición o extensión de estas zonas grises, ya que obliga a las partes a intercambiar información sobre aspectos de los sistemas de inteligencia artificial que puedan tener un impacto, ya sea positivo o negativo, en la protección de los derechos humanos, la democracia o el Estado de derecho. Esta cooperación también se prevé en el ámbito de la investigación, la evaluación de los riesgos y en la interlocución con el sector privado (art. 25). En efecto, aunque las empresas no puedan ser parte de la Convención Marco —al no ser sujetos de derecho internacional—, sí se contempla su implicación. Su participación resulta imprescindible, dada la naturaleza de la materia, y por ello el tratado prevé mecanismos de cooperación con el sector privado, reconociendo su papel clave en el desarrollo y despliegue de los sistemas de inteligencia artificial.¹²

De la interacción entre la Convención Marco y el Reglamento de inteligencia artificial aún queda por ver si prevalecerá el denominado “efecto Bruselas” —esto es, la capacidad de la normativa de la Unión Europea para liderar y establecer estándares globales en la materia— o, por el contrario, que las partes que no son miembros impulsen otras direcciones. ¿Quién influirá a quién en esta tensión por definir el ritmo regulatorio de la inteligencia artificial? A día de hoy la respuesta sigue siendo incierta (Colomina y Galceran-Vercher, 2024).

¹² Este modelo de participación privada no se comparte en todos los países del mundo. China, los Estados del Golfo o Rusia apuestan por un modelo más estatista de la gestión de los temas digitales sin participación de actores no estatales (Galcerán, 2024).

2.4. Las resoluciones de Naciones Unidas sobre inteligencia artificial

En el ámbito internacional de alcance universal, es decir, en el marco de las Naciones Unidas, apenas existen textos con valor jurídico vinculante que orienten la regulación de la inteligencia artificial; tan solo se han emitido documentos de carácter orientativo que señalan posibles líneas de actuación. La Resolución de la Asamblea General de las Naciones Unidas (AGNU) sobre IA, de 11 de marzo de 2024¹³ promueve la cooperación internacional y alienta a los Estados a que dispongan de sistemas seguros, protegidos y fiables de IA utilizando, entre otros medios, la transparencia. Sin embargo, el texto no ofrece una definición clara de lo que se entiende por transparencia. No obstante, vincula la transparencia con la previsibilidad, la fiabilidad y la comprensibilidad a lo largo de todo el ciclo de vida de los sistemas de IA utilizados para tomar o respaldar decisiones que afecten a los usuarios finales. En otras palabras, la AGNU entiende que la transparencia resulta aplicable en la medida en que los sistemas de IA puedan tener un impacto directo sobre las personas usuarias. En este sentido, los aspectos que debería abarcar la transparencia incluyen:

- dar información y explicaciones a los usuarios finales,
- promover la supervisión humana,
- rendir cuentas eficaces para los que se vean afectados negativamente por una decisión adoptada de forma autónoma por la IA (punto 6.k).

Igualmente, aunque el redactado es bastante confuso, en la Resolución se alienta a los Estados miembros a que protejan los datos personales al poner a prueba y evaluar la IA y a que se tenga en cuenta dicha protección, cuando haya que presentar informes sobre el uso de esos datos de conformidad con las legislaciones nacionales e internacionales (aquí se podría incluir la europea), bajo el requisito de la transparencia (punto 6.j).

Por otro lado, en el marco también de la AGNU hay que citar la resolución sobre el *Pacto para el Futuro* que fue adoptado el 22 de septiembre de 2024.¹⁴ En ésta se incluye un Anexo I dedicado al “Pacto digital global” que introduce un punto dedicado a la mejora de la gobernanza internacional de la IA en beneficio de la humanidad. Así, dentro de este objetivo, que vuelve a situar el bienestar de las personas en el centro de atención, en consonancia con la legislación europea, la transparencia es considerada un requisito clave para lograrlo. En particular, insta a la cooperación internacional para promover las tecnologías digitales, en las que incluye la IA, basada en la responsabilidad, la rendición de cuentas y la transparencia, destacando este último requisito de forma explícita (punto 55.d). De esta manera, los Estados miembros se comprometen a “promover la transparencia” pero sin más explicaciones.¹⁵

¹³ AGNU; “Aprovechar las oportunidades de sistemas seguros, protegidos y fiables de inteligencia artificial para el desarrollo sostenible”, A/78/L.49, 11 de marzo de 2024

¹⁴ AGNU; “El Pacto para el Futuro”, A/RES/79/1, 22 de septiembre de 2024

¹⁵ En la resolución de la AGNU sobre el Pacto digital global, se asume el compromiso de hacer un seguimiento de la misma, como organizar una reunión denominada “Examen de alto nivel” que se basará

En suma, como se ha venido observando, prevalece una notable imprecisión respecto al valor que debe otorgarse al principio de transparencia en los sistemas de inteligencia artificial, tanto a nivel europeo como internacional.

3. OBLIGACIONES DE TRANSPARENCIA SEGÚN EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL

El propio Reglamento de inteligencia artificial constituye un ejemplo claro de cómo, dentro de una misma norma, el principio de transparencia puede abordarse de distintas formas y con diferentes alcances. En concreto, prevé obligaciones de transparencia para proveedores y responsables del despliegue que varían en función del mayor o menor riesgo de los sistemas de IA. De este modo, el artículo 50 establece obligaciones de transparencia para “determinados sistemas de IA”, dejando fuera a los sistemas de IA de alto riesgo. En efecto, el Reglamento distingue, respecto de las obligaciones de transparencia, entre los sistemas de IA de alto riesgo, que tienen que cumplir con muchos más requisitos, que el resto (art. 50.6). Este artículo no aborda la transparencia en los sistemas de IA de alto riesgo, pero conviene tener presente que las exigencias en este ámbito son mayores.¹⁶ Así, el artículo 50 distingue entre proveedores y responsables del despliegue a efectos de transparencia.

3.1. Proveedores

3.1.1. Artículo 50.1 del Reglamento de Inteligencia Artificial: Obligación de informar en interacciones directas con personas físicas

Los proveedores de sistemas de inteligencia artificial que interactúan directamente con personas físicas están obligados a informar de dicha interacción. El artículo 50.1 del Reglamento de inteligencia artificial impone una clara obligación de transparencia, según la cual toda persona debe ser advertida de forma explícita cuando está interactuando con un sistema de IA.

Esta obligación no se aplicará a los sistemas de IA legalmente autorizados para la detección, prevención, investigación o enjuiciamiento de delitos penales, siempre que su uso no implique una interacción directa y pública con personas físicas.

No obstante, existe una excepción a esta excepción: Cuando el sistema de IA esté disponible públicamente con el fin de permitir la denuncia de un delito penal, entonces sí se mantiene la obligación de informar a la persona usuaria.

en un informe del Secretario General sobre los progresos alcanzados. Es posible que en dichas reuniones se aborde el tema de la transparencia en el ámbito de la IA dado que es un aspecto más que aparece vinculado a ésta.

¹⁶ En este artículo no se abordan las obligaciones de transparencia para sistemas de IA de alto riesgo cuyas exigencias son mayores que las previstas para otros modelos. Ello requerirá un nuevo trabajo, una segunda parte, que parte de este primer artículo como marco referencial de la transparencia.

3.1.2. Artículo 50.1 del Reglamento de Inteligencia Artificial: Obligación de informar en interacciones con personas físicas

Los proveedores de sistemas de inteligencia artificial que interactúan directamente con personas físicas están obligados a informarles expresamente de dicha interacción. Así lo dispone el artículo 50.1 del Reglamento de inteligencia artificial, que establece una obligación de transparencia: toda persona debe ser claramente advertida cuando está interactuando con un sistema de IA, a fin de garantizar un uso consciente y responsable de esta tecnología.

Esta obligación no será exigible cuando se trate de sistemas de IA legalmente autorizados para detectar, prevenir, investigar o enjuiciar delitos penales, siempre que su utilización no implique una interacción directa y pública con personas físicas.

No obstante, esta excepción no será aplicable cuando el sistema esté disponible al público para facilitar la denuncia de delitos penales. En ese caso, sí se mantiene la obligación de informar al usuario de que está interactuando con una inteligencia artificial.

3.1.3. Artículo 50.1 del Reglamento de Inteligencia Artificial: Obligación de información en interacciones con personas físicas

Los proveedores de sistemas de inteligencia artificial que interactúan directamente con personas físicas están obligados a informar expresamente de ello. El artículo 50.1 del Reglamento de inteligencia artificial establece una obligación de transparencia, según la cual toda persona debe ser claramente advertida cuando está interactuando con un sistema de IA. Esta medida busca garantizar que las personas usuarias sean conscientes de que no están tratando con otro ser humano, sino con una entidad automatizada.

Esta obligación no será exigible en los casos en que el sistema de IA esté legalmente autorizado para detectar, prevenir, investigar o enjuiciar delitos penales, siempre que su uso no implique una interacción directa y pública con individuos.

Sin embargo, existe una excepción a esta excepción: Cuando el sistema esté disponible al público con el propósito de facilitar la denuncia de un delito penal, sí deberá mantenerse la obligación de informar a la persona usuaria de que está interactuando con inteligencia artificial.

3.1.4. Artículo 50.1 del Reglamento de Inteligencia Artificial: Obligación de información en interacciones con personas físicas

Los proveedores de sistemas de inteligencia artificial que interactúan directamente con personas físicas están obligados a informar claramente de dicha interacción. El artículo 50.1 del Reglamento de inteligencia artificial impone una obligación de transparencia, conforme a la cual toda persona debe ser advertida de forma explícita

cuando está interactuando con un sistema de IA, con el objetivo de garantizar decisiones informadas y proteger los derechos de los usuarios.

Esta obligación no se aplicará cuando se trate de sistemas de IA legalmente autorizados para detectar, prevenir, investigar o enjuiciar delitos penales, siempre que su uso no implique una interacción directa y pública con personas físicas.

No obstante, existe una excepción a esta excepción: Si el sistema de IA está disponible al público para facilitar la denuncia de delitos penales, entonces sí debe mantenerse la obligación de informar al usuario de que está interactuando con inteligencia artificial.

3.1.5. Artículo 50.2 del Reglamento de Inteligencia Artificial: Obligaciones de legibilidad, información y fiabilidad en contenidos sintéticos

Los proveedores de sistemas de inteligencia artificial —incluidos los de uso general¹⁷— que generen contenido sintético (imagen, audio, texto o vídeo), deberán garantizar que los resultados producidos por dichos sistemas estén marcados en un formato legible por máquina y que permitan identificar de forma clara que han sido generados o manipulados artificialmente. Se trata, por tanto, de obligaciones de legibilidad y de información, orientadas a prevenir la confusión entre contenidos reales y artificiales (art. 50.2).

Además, estos proveedores están obligados a velar por que las soluciones técnicas que implementen sean interoperables, eficaces, sólidas y fiables, en la medida en que ello sea técnicamente posible. Esta disposición introduce una obligación de fiabilidad, que se extiende al diseño, funcionamiento y comunicación de los sistemas de IA. Estas obligaciones no se aplicarán en los siguientes supuestos:

- Cuando el sistema de IA desempeñe una función de apoyo a la edición estándar.
- Cuando no altere sustancialmente los datos de entrada proporcionados por el responsable del despliegue ni su semántica.
- O cuando el sistema esté legalmente autorizado para la detección, prevención, investigación o enjuiciamiento de delitos (art. 50.2).

Los redactores del Reglamento eran plenamente conscientes de la presión existente para sacarlo adelante, ya que la Unión Europea aspiraba a liderar la legislación en esta materia. No obstante, esta ambición se enfrentaba a la dificultad de no contar aún con todas las claves ni certezas sobre los posibles problemas —incluidos los de orden jurídico— que podrían surgir con la aplicación real de los sistemas de

¹⁷ Un sistema de IA de uso general, según el Reglamento: “un sistema de IA basado en un modelo de IA de uso general y que puede servir para diversos fines, tanto para su uso directo como para su integración en otros sistemas de IA” (art. 3. 66). Por ejemplo, *ChatGPT* de *OpenAI* podría ser considerado un sistema de IA de uso general. <https://www.europarl.europa.eu/topics/es/article/20230601STO93804/ley-de-ia-de-la-ue-primera-normativa-sobre-inteligencia-artificial>

inteligencia artificial. Los proveedores y los responsables del despliegue de sistemas de inteligencia artificial se enfrentan a notables dificultades a la hora de cumplir con el Reglamento, especialmente debido a la falta de claridad sobre la documentación concreta que deben presentar para satisfacer sus obligaciones en materia de transparencia. Por este motivo, la elaboración de códigos de prácticas a escala de la Unión, por parte de la Oficina de la IA, resulta imprescindible para facilitar la aplicación efectiva de dichas obligaciones.

Además, la propia evolución de los sistemas de IA —cuyas consecuencias pueden ser, en ocasiones, impredecibles— exige que estos códigos de buenas prácticas se mantengan en constante actualización. En este sentido, cabe valorar positivamente que el legislador haya previsto esta necesidad y establecido un procedimiento específico para ello, tal como se recoge en el artículo 50.7 del Reglamento.

En este sentido, la Oficina de Inteligencia Artificial ha elaborado un primer y un segundo borrador del código de prácticas relativo a los sistemas de IA de uso general. El segundo borrador incorpora las observaciones y comentarios formulados por las partes interesadas en relación con la versión inicial.¹⁸ El primero desarrolla algunos de los puntos del artículo 53, que están relacionados con la transparencia, para una mayor comprensión de quienes tengan que cumplir con dichos requisitos y proporcionarles seguridad de que están actuando conforme a la ley. Así, elabora dos columnas: una destinada a las obligaciones que tienen que cumplir los proveedores respecto de la Oficina de IA y autoridades nacionales competentes y, otra, respecto a proveedores intermedios. Ahora bien, el artículo 53 del Reglamento, que ha sido objeto de desarrollo en el primer borrador de código de prácticas, en ningún momento se menciona la transparencia. Se puede deducir, efectivamente, pero hubiera sido deseable que se incluyera alguna mención como sí hizo en el artículo 50. El borrador de código de prácticas suple dicha omisión al abordar los requerimientos del artículo 53 dentro del apartado que han denominado “Transparencia”.

Por su parte, el segundo borrador cuyo contenido será sometido también a discusión, introduce nuevos elementos. En particular, señala que determinadas medidas de transparencia no se aplicarán a los proveedores de modelos de IA de código abierto que cumplan las condiciones especificadas en el artículo 53.2 del Reglamento, a menos que tales modelos de IA de uso general sean considerados con riesgo sistémico. En cuanto a la posibilidad de los que suscriban el Código de prácticas de informar a los usuarios, en todo o en parte, de las informaciones que proporcionan con el objetivo de promover la transparencia pública que estaba contemplado en el primer borrador, en este segundo se añade que una parte de dicha información también podrá solicitarse para dar cumplimiento al art. 53. 1. d. Es decir, deberán elaborar y poner a disposición del público un resumen del contenido utilizado para el entrenamiento del modelo de IA de uso general, con arreglo al modelo facilitado por la Oficina de IA, que

¹⁸ Véase los comentarios en la siguiente dirección:

https://administracionelectronica.gob.es/pae_Home/pae_Actualidad/pae_Noticias/2024/Diciembre/noticia-2024-12-20-Comite-Europeo-IA-publica-segundo-borrador-codigo-practicas-IA-.html

es una plantilla que está en proceso de elaboración. Dado que estos documentos aún se encuentran en proceso de revisión, cabe destacar que su finalidad principal es alcanzar un consenso, conforme a lo previsto en el artículo 53 del Reglamento, sobre qué documentación debe proporcionarse para dar cumplimiento a las obligaciones de transparencia.

Para más detalles y dado que los dos borradores de código de prácticas están en *Adobe Acrobat y Reader*, se ha preguntado al Asistente de IA que se estrenó el pasado diciembre de 2024 para los documentos en formato PDF, en un ejercicio experimental, acerca de las diferencias entre el primer y el segundo borrador respecto de la transparencia. Estas son sus respuestas:

1. *Claridad y detalle.* El segundo borrador proporciona más detalles y ejemplos concretos sobre los compromisos de transparencia. Se han añadido tablas y descripciones específicas para cada tipo de información que los proveedores deben documentar y mantener actualizada.
2. *Compromiso de documentación.* Se ha especificado que los signatarios deben considerar si la información documentada puede ser divulgada al público para promover la transparencia pública. También se ha incluido un compromiso para asegurar la calidad, integridad y seguridad de la información documentada.
3. *Elementos esenciales de una Política de Uso Aceptable.* Se ha añadido un apéndice que detalla los elementos esenciales que debe contener una política de uso aceptable, proporcionando una guía más clara para los proveedores.
4. *Indicadores Clave de Desempeño (KPIs).* Se han introducido KPIs preliminares para medir la efectividad de los compromisos de transparencia, lo que no estaba presente en el primer borrador.
5. *Proceso de Evaluación y Mitigación de Riesgos.* Se ha detallado más el proceso de evaluación y mitigación de riesgos, incluyendo la necesidad de realizar evaluaciones periódicas y la documentación de los resultados de estas evaluaciones.
6. *Notificación y Publicación.* Se ha especificado que los signatarios deben notificar a la Oficina de IA sobre la clasificación de sus modelos y proporcionar acceso a la versión más reciente de su Marco de Seguridad y Reportes de Modelos. También se ha aclarado que los informes de modelos deben ser enviados a la Oficina de IA al momento de colocar un modelo en el mercado.

Estas diferencias reflejan un esfuerzo por hacer el código más detallado, específico y alineado con las mejores prácticas y estándares internacionales, así como por proporcionar una mayor claridad y guía a los proveedores de modelos de IA.

La comprobación (humana) que en este caso se realiza para valorar los comentarios del Asistente de IA es positiva: la sistematización es clara, aborda en extenso todos los puntos, y responde en segundos.

3.2. Responsables del despliegue

3.2.1. Artículo 50.3 del Reglamento de Inteligencia Artificial: Obligaciones relativas al despliegue de sistemas de reconocimiento de emociones y categorización biométrica

En relación con los responsables del despliegue de sistemas de reconocimiento de emociones o de categorización biométrica, el artículo 50.3 del Reglamento de Inteligencia Artificial establece que:

- deben informar a las personas físicas expuestas al funcionamiento del sistema, y
- deben tratar sus datos personales de conformidad con la normativa europea en materia de protección de datos, especialmente con el Reglamento General de Protección de Datos (RGPD)

Se trata, por tanto, de una doble obligación:

- Obligación de información, para garantizar la transparencia ante los individuos afectados, y
- obligación de protección de los datos personales, alineada con los principios de licitud, lealtad y minimización del RGPD.

Esta obligación no se aplica a los sistemas de IA utilizados para la categorización biométrica o el reconocimiento de emociones que hayan sido autorizados por ley para detectar, prevenir o investigar delitos, siempre que dichas actividades se desarrollen de conformidad con el Derecho de la Unión Europea (art. 50.3).

3.2.2. Artículo 50.4 del Reglamento de IA: Responsabilidad en casos de ultrasuplantación (deepfakes)

Los responsables del despliegue de un sistema de inteligencia artificial que genere o manipule imágenes, audio o vídeo de forma que constituya una ultrasuplantación (deepfake) estarán obligados a:

- hacer público que dichos contenidos han sido generados o manipulados artificialmente;
- garantizar que esta información esté accesible y claramente identificable para el público.

Se trata de una obligación de publicidad, orientada a preservar la confianza, la veracidad informativa y la protección frente a usos engañosos o maliciosos de la IA.

Cuando el contenido forme parte de una obra manifiestamente creativa, satírica, artística, de ficción o análoga, la obligación se matiza:

- Se mantiene la obligación de hacer pública la existencia del contenido generado o manipulado artificialmente,
- pero esta debe aplicarse de manera que no dificulte la exhibición ni el disfrute de la obra.

En este supuesto, la obligación de publicidad se mantiene, pero se limita en su forma, para no interferir con la integridad estética, narrativa o funcional de la obra creativa.

Esta obligación no se aplicará cuando el uso de dichos sistemas esté autorizado por ley para detectar, prevenir, investigar o enjuiciar delitos, de acuerdo con el marco legal de la Unión Europea.

3.2.3. Artículo 50.4 del Reglamento de Inteligencia Artificial: Obligación de divulgación en contenidos informativos generados por IA

Los responsables del despliegue de un sistema de inteligencia artificial que genere o manipule texto con el propósito de informar al público sobre asuntos de interés general estarán obligados a divulgar que dicho contenido ha sido generado o manipulado de forma artificial.

Se trata de una obligación de divulgación, orientada a garantizar la transparencia informativa y a prevenir la desinformación en contextos donde los contenidos generados por IA puedan influir en la opinión pública o en decisiones de relevancia social. Esta obligación no se aplicará en los siguientes supuestos:

- cuando el uso del sistema esté autorizado por ley para la detección, prevención, investigación o enjuiciamiento de delitos;
- cuando el contenido generado por IA haya sido sometido a un proceso de revisión humana o control editorial;
- o cuando exista una persona física o jurídica que asuma la responsabilidad editorial de la publicación del contenido.

3.2.4. Artículo 50.6 del Reglamento de Inteligencia Artificial: Aplicación complementaria de obligaciones de transparencia

En relación con los responsables del despliegue de sistemas de inteligencia artificial, el Reglamento establece que el conjunto de obligaciones vinculadas a la transparencia no se agota en su propio articulado.

Según el artículo 50.6, también serán aplicables aquellas obligaciones adicionales que se deriven del Derecho nacional o del Derecho de la Unión Europea.

Esto implica que los responsables deben tener en cuenta no solo el marco específico del Reglamento de IA, sino también otras normas sectoriales, nacionales o comunitarias que puedan imponer requisitos adicionales en materia de información, trazabilidad, supervisión o rendición de cuentas.

A modo de cierre de este apartado y respecto a las exigencias de cumplimiento de las obligaciones de transparencia por parte de las personas físicas, el Reglamento prevé que la información que se les ha de proporcionar ha de efectuarse de forma clara y distinguible. Por tanto, las obligaciones de informar y divulgar han de cumplir con los requisitos de claridad y distinguibilidad. Asimismo, proveedores y responsables del despliegue tienen una exigencia de temporalidad: la información se ha de proporcionar a más tardar en la primera interacción o exposición (art. 50.5). En este sentido, la exigencia de transparencia implica que las personas físicas han de saber que están interactuando con un sistema de IA desde el primer momento.

4. TRANSPARENCIA E INCERTEZAS

4.1. Los contornos de la transparencia

A pesar de la entrada en vigor del Reglamento de inteligencia artificial y de la adopción de una Convención Marco en el seno del Consejo de Europa —y de que ambos instrumentos abordan la cuestión de la transparencia en los sistemas de IA—, su conceptualización, alcance y efectos prácticos siguen siendo inciertos.

El Reglamento no ofrece una definición precisa del concepto de transparencia. Si bien dedica varios artículos a regular obligaciones relacionadas con ella, en otros —como el artículo 53— ni siquiera se la menciona expresamente. No obstante, dichos preceptos sí entran dentro de su ámbito de aplicación, tal como ha quedado reflejado en los borradores del código de prácticas elaborados hasta el momento. A su vez, remite a informes no vinculantes para establecer qué entra dentro de la transparencia, o remite a otras normas europeas como el Reglamento europeo de protección de datos. Este último califica la transparencia como un principio **de** valor vinculante, al igual que lo hace la Convención Marco sobre Inteligencia Artificial, la cual establece que, una vez entre en vigor, la transparencia será un principio obligatorio aplicable a todas las actividades desarrolladas a lo largo del ciclo de vida de los sistemas de IA. Las resoluciones de la Asamblea General de las Naciones Unidas en esta materia tampoco aportan claridad sobre esta cuestión. Su valor reside, más bien, en reafirmar la importancia de la transparencia como un requisito fundamental que debe guiar la regulación de los sistemas de inteligencia artificial a nivel nacional, europeo e internacional a lo largo de todo su ciclo de vida, pero sin entrar en el fondo.

Así, la falta de armonización en la definición y alcance de los conceptos dificulta una aproximación más eficaz del principio de transparencia. A modo de ejemplo, la transparencia según el Reglamento General de Protección de Datos (RGPD), de aplicación nacional, difiere de la contemplada en el Reglamento de inteligencia artificial

en aspectos clave como los actores implicados, el tipo de información requerida y los destinatarios a los que va dirigida.¹⁹

En cuanto al contenido, no queda claro si la transparencia constituye un concepto independiente o si, por el contrario, la explicabilidad y la trazabilidad son simplemente dimensiones que se derivan de ella. Esta ambigüedad conceptual genera una notable confusión. Mientras que el Reglamento de Inteligencia Artificial, apoyado en un informe del grupo de expertos, considera que la transparencia engloba elementos como la explicabilidad, la trazabilidad y la comunicación; la Convención Marco sobre la Inteligencia Artificial, a través de su Informe Explicativo, amplía esta visión incluyendo también la accesibilidad y la interpretabilidad.

En este contexto, la explicabilidad se entiende como la capacidad de ofrecer razones comprensibles sobre por qué un sistema de IA toma una determinada decisión o realiza una recomendación. Por su parte, la interpretabilidad hace referencia a la posibilidad de comprender el funcionamiento interno del sistema, es decir, cómo toma dichas decisiones.

Respecto a la comunicación, esta implica informar de manera clara y accesible sobre aspectos fundamentales como el origen de los datos, los procesos de entrenamiento del sistema, las pruebas realizadas, las validaciones efectuadas y las fuentes utilizadas.

En los textos analizados, la explicabilidad aparece de forma consistente como un componente esencial de la transparencia. Por su parte, las resoluciones de la Asamblea General de las Naciones Unidas (AGNU) también hacen referencia —aunque de forma implícita y sin nombrarlas expresamente— a dimensiones como la explicabilidad y la comunicación, lo que refuerza su relevancia dentro del principio general de transparencia.

En cambio, el reciente Libro Blanco sobre Transparencia de los sistemas de IA de la *Federal Office for Information Security* de Alemania²⁰ considera las nociones de trazabilidad y explicabilidad distintas a la transparencia. Esta última la asocian con la confiabilidad, y el propio Libro blanco admite (p, 5) que los distintos elementos no pueden diferenciarse claramente entre sí, pero que cada uno se centra en zonas temáticas diversas. Estas incertezas persisten debido a la ausencia de un concepto consensuado de transparencia en el ámbito de los sistemas de inteligencia artificial, lo que provoca variaciones significativas en las características que se le atribuyen. No obstante, a pesar de esta falta de armonización, pueden extraerse algunas consideraciones comunes que permiten esbozar una comprensión más estructurada del principio:

Coincidencias. Primero, la transparencia ha de estar presente durante todo el ciclo de vida de los sistemas de IA, desde el diseño, desarrollo y validación, puesta en

¹⁹ AEPD (Agencia Española de Protección de Datos); “Inteligencia artificial: transparencia”, <https://www.aepd.es/prensa-y-comunicacion/blog/inteligencia-artificial-transparencia>, 20 de septiembre de 2023

²⁰ Federal Office for Information Security (2024). *Transparency of AI Systems. White paper.*

servicio y aplicación e igualmente en las actualizaciones del sistema. Segundo, la exigencia de transparencia va dirigida a proteger a las personas frente a los riesgos de los sistemas de IA.

Beneficios. El Libro blanco (p. 13-14) apunta a que la transparencia promueve la trazabilidad de las decisiones y la evaluación sobre su conveniencia. Protege contra el abuso al permitir que mediante la transparencia se minimicen los riesgos y daños contra las personas en fases tempranas. Mediante la transparencia se puede evaluar un sistema de IA y ello abunda en la confiabilidad. La transparencia detecta vulnerabilidades del sistema (si se hace todo en oculto no se ve). Asimismo, según el Libro blanco (p. 13), favorece el empoderamiento de las partes interesadas y fortalece la confiabilidad del usuario en el sistema de IA. Sirva de ejemplo, y como ya se comentó, *Adobe Acrobat Reader* dispone de un “Asistente de IA” al que se le puede preguntar cualquier cuestión relacionada con el documento que el usuario abre. El mismo asistente te advierte que su uso puede dar problemas de fiabilidad:

Adobe protege tus documentos y su contenido con prácticas responsables. Comprueba las respuestas y las fuentes, ya que es posible que no siempre sean precisas. No uses el Asistente de IA para obtener asesoramiento profesional.

El propio asistente de IA remite a unas directrices recientemente aprobadas por Adobe,²¹ que regulan de forma más desarrollada su uso. Si bien este enfoque puede resultar farragoso desde el punto de vista operativo, contribuye a reforzar la confianza en que se trata de un sistema de inteligencia artificial susceptible de ser manejado y controlado por el usuario.

4.2. Los datos para entrenar los sistemas de IA

Otro aspecto muy relevante y objeto de disputa es en relación con los datos. El Libro blanco enfatiza (p. 5) que la información sobre los datos de entrenamiento es esencial para evaluar el riesgo de ataques de envenenamiento del sistema. La constatación de este hecho requiere que los sistemas de IA a lo largo de todo su ciclo vital permitan la trazabilidad y explicabilidad y como señala el propio Libro Blanco deben ir acompañados del criterio de transparencia que se integra en el campo de la confiabilidad de los sistemas de IA.

En contraste, Martin Coulter (2024) advierte que exigir una mayor transparencia a las compañías en relación con los datos utilizados para entrenar sus sistemas podría forzarlas a revelar uno de sus secretos mejor guardados. En este sentido, si se vieran obligadas, por ejemplo, a informar sobre el uso de películas, libros superventas u otras obras protegidas que han empleado sin el consentimiento de sus autores, se abriría inevitablemente una brecha con los derechos de propiedad

²¹ Directrices para usuarios de la IA generativa de Adobe. <https://www.adobe.com/es/legal/licenses-terms/adobe-gen-ai-user-guidelines.html>

intelectual. Además, como señala el propio autor, hacer pública esta información también supone ofrecer una ventaja competitiva injusta a otras empresas del sector. Ante esta situación, cabe preguntarse qué implicaciones traerá consigo el cumplimiento de las obligaciones de transparencia tanto para las pequeñas *startups* dedicadas a la inteligencia artificial, como para las grandes corporaciones tecnológicas —como *Google* o *Meta*—, cuyas estrategias de crecimiento se apoyan firmemente en el desarrollo de avances tecnológicos.

Ya hay ejemplos de demandas presentadas por los creadores contra grandes empresas —como *Google* o *OpenAI*— por haber utilizado impropriamente contenidos protegidos por derechos de autor para entrenar sus modelos. Y aunque se está llegando a acuerdos de licencias para usar dichos contenidos, no está todo resuelto. Martin Coulter ilustra esta tensión con el caso de las críticas dirigidas a *OpenAI*, cuando, la entonces *Chief Technology Officer (CTO)* de la compañía, Mira Murati, se negó a responder a una pregunta del *Wall Street Journal* sobre si se habían utilizado vídeos de YouTube para entrenar su herramienta generativa de vídeo, conocida como *Sora*.²² Esta IA, abierta al público en general, genera contenidos audiovisuales de forma artificial. O cuando *OpenAI* utilizó una voz generada por la IA muy parecida a la de la actriz Scarlett Johansson para lanzar una versión más nueva de su *ChatGPT*.

Por si fuera poco, en septiembre de 2024, varios directores ejecutivos, analistas y estrategas de reconocido prestigio en el ámbito de la inteligencia artificial —procedentes de entidades como *Spotify*, *Meta*, *Ericsson*, el *Tony Blair Institute* o la *Harvard University*, entre otros— firmaron una carta abierta titulada significativamente: *Europe needs regulatory certainty on AI* (Ericsson, 2024). En dicha publicación, los firmantes expresan su preocupación por la inseguridad normativa existente en relación con el uso de datos para el entrenamiento de modelos de inteligencia artificial. Señalan, en particular, que las actuaciones de diversas autoridades europeas de protección de datos están generando una profunda incertidumbre sobre qué tipos de datos pueden utilizarse legalmente, lo que dificulta la planificación y el desarrollo de proyectos tecnológicos a gran escala dentro del marco europeo. Asimismo, los firmantes advierten que esta situación podría provocar que los modelos de inteligencia artificial de código abierto terminen desarrollándose sin tener en cuenta la cultura, los valores y el conocimiento europeo —dándose a entender, de forma implícita, que ello se debe a la exclusión de datos procedentes de la Unión Europea. En consecuencia, reclaman una interpretación más actualizada y flexible del Reglamento General de Protección de Datos (RGPD) que sea compatible con la necesidad de utilizar datos personales de ciudadanos europeos para el entrenamiento de modelos de IA. A su juicio, esta adaptación normativa resultaría esencial para evitar

²² Para entrar más en detalles sobre la entrevista de Mira Murati en relación con *Sora* y la falta de respuestas en cuestiones relacionadas con la transparencia. LLM Máster; “*Sora OpenAI - Mira Murati CTO de OpenAI confirma el lanzamiento de Sora al público*”, 15 de marzo de 2024, <https://youtu.be/IS0G2D6MKGw?si=htYD4aM1V1-vzjRC>

que Europa quede rezagada en términos de competitividad e innovación tecnológica en el ámbito de la inteligencia artificial.

Luiza Jarovsky (2024), por el contrario, considera que la normativa europea sí que es segura. El propio Reglamento General de Protección de Datos (RGPD) contempla la posibilidad de que las personas otorguen —o no— su consentimiento expreso para que sus datos sean utilizados en el entrenamiento de sistemas de inteligencia artificial. Esta vía permitiría, al menos en teoría, sortear el obstáculo que representa la imposibilidad de utilizar datos procedentes de la Unión Europea, ofreciendo así una solución jurídica viable que compatibiliza el respeto a los derechos fundamentales con el desarrollo tecnológico. Sin embargo, como Jarovsky sostiene, la mayoría de las compañías no quieren ser transparentes con los datos que usan para entrenar sus IA. Menciona el caso de *Meta*, en que después de múltiples quejas sobre el uso de datos personales, la compañía decidió, en un primer momento, en lugar de preguntar a los usuarios si cedían sus datos para entrenarla, no usar datos personales de la UE. Está claro que lo ideal sería que hubiera un equilibrio entre los derechos de autor y la necesidad de que sus obras sirvan para el entreno de los modelos de IA. Habrá que ver, una vez que sea completa su aplicación, si el Reglamento garantiza tanto la protección de los secretos comerciales como los derechos de los autores. Muy recientemente, la entrada al mercado de la IA de fabricación china *DeepSeek*, con un coste mucho menor que otros modelos como *ChatGPT*, ha dado un vuelco al mercado de las IA. Las autoridades italianas ya han pedido explicaciones sobre cómo *DeepSeek* se entrena y qué datos personales se están utilizando ²³ (El Tiempo & Infofae, 2025).

4.3. Los algoritmos y sus sesgos

Otra cuestión de gran relevancia en el debate actual es si la transparencia debe extenderse también a los propios algoritmos utilizados por los sistemas de inteligencia artificial. Alberto Pena, del *European Centre for Algorithmic Transparency*,²⁴ señalaba en una entrevista publicada en el Diario La Ley (2024) una reflexión que resulta perfectamente aplicable a los algoritmos utilizados en los sistemas de inteligencia artificial:

cualquier usuario de una plataforma digital debe saber qué puede esperar de ella, debe poder predecir la respuesta que le van a ofrecer, para tener la última palabra al respecto. Se trata, esencialmente, de evitar la opacidad de las

²³ Entre otros: Italia pide explicaciones a DeepSeek sobre la utilización de datos personales en su nueva IA. <https://www.eltiempo.com/mundo/europa/italia-pide-explicaciones-a-deepseek-sobre-la-utilizacion-de-datos-personales-en-su-nueva-ia-3421903>
<https://www.infobae.com/tecno/2025/01/29/deepseek-desaparece-de-italia-por-preocupaciones-por-la-privacidad-de-los-datos/>

²⁴ El Centro Europeo para la Transparencia Algorítmica (ECAT) se puso en marcha en abril de 2023 para proporcionar experiencia científica y técnica para respaldar la aplicación de la Ley de Servicios Digitales (DSA) y promover la investigación sobre el impacto de los sistemas algorítmicos implementados por plataformas en línea y motores de búsqueda.

denominadas cajas negras algorítmicas, cuyo funcionamiento no se puede conocer.

¿Hay que preocuparse por los sesgos algorítmicos? Es evidente que los algoritmos pueden tener sesgos en los datos, en las mediciones, en el modelo, en la evaluación, en la información aportada, en la selección, entre otros (Pathak, 2024). Susana Navas (2023) explica de forma exhaustiva las causas de los sesgos en los algoritmos, que no son pocas. En este sentido, la transparencia sobre el algoritmo promueve la confianza entre los usuarios de la IA al entender cómo funciona la IA y los resultados que produce. Estas cualidades estarían asociadas a la explicabilidad e interpretabilidad.

Carlos Fernández y Josu Andoni Eguiluz (2024) señalan como un punto crítico del Reglamento la escasa atención que se dedica al principio de explicabilidad, el cual consideran estrechamente vinculado con la transparencia. Subrayan que esta carencia resulta especialmente preocupante cuando se trata de algoritmos, ya que la explicabilidad permite comprender no solo el resultado que produce un sistema, sino también el proceso mediante el cual se ha alcanzado dicho resultado.

Ignacio G.R Gavilán (2022) razona acerca de las características que ha de tener la explicabilidad: explicación, significado, precisión, límites, a partir del estudio de textos comparados, y llega a la conclusión que no siempre va a poder aplicarse dicho principio por sus dificultades técnicas. Sin embargo, como también señala Ignacio G.R Gavilán (2024):

si los algoritmos 'razonan' o calculan de una forma algorítmico-matemática, una forma que no es la humana, conseguir, si es que lo consiguiéramos, que nos explicaran cómo han razonado en términos simbólico-lingüísticos, en términos humanos, podría ser, en el fondo, sólo una aproximación, una ayuda, casi una metáfora, quizá incluso una impostura, pero nunca una verdadera explicación de cómo han llegado a sus conclusiones.

Incluso llega a plantearse que quizás en el caso de los algoritmos considerados de *deep learning*, la explicabilidad puede ser un problema irresoluble, dadas las propias características opacas de estos modelos.

5. CONCLUSIONES: RIESGOS POR EXCESO O DEFECTO DE TRANSPARENCIA

Revelar según qué información sobre el diseño de un sistema de inteligencia artificial facilita su exposición a ciberataques. Los hackers aprovecharían este conocimiento para identificar vulnerabilidades y explotarlas. Además, dicha información podría ser utilizada para inducir comportamientos erróneos o dañinos en el sistema, por ejemplo, mediante técnicas como el *envenenamiento de datos*.

También existe el riesgo de que, bajo el amparo de obligaciones de transparencia, se proporcione intencionadamente información errónea, lo que podría distorsionar la evaluación del sistema por parte de usuarios, auditores o reguladores.

En este sentido, ¿Puede aplicarse el principio de economía de datos del Reglamento General de Protección de Datos (RGPD) de la Unión Europea como criterio para abordar este problema? El Libro Blanco sobre la Inteligencia Artificial así lo sugiere: las partes interesadas deben acceder únicamente a la información estrictamente necesaria para poder tomar una decisión informada, y no más.

Vadim Fedorov (2024), experto en desarrollo de *software*, advierte que las actuales obligaciones de transparencia resultan insuficientes frente a la complejidad de los sistemas de inteligencia artificial. No basta con indicar que se está utilizando un sistema de inteligencia artificial; también es necesario advertir que los resultados que este proporciona pueden ser erróneos, y que tales errores pueden tener un impacto directo en la vida de las personas.

El autor ilustra esta idea con un paralelismo muy revelador: la ley de *cookies*. En la práctica, cuando un usuario accede a cualquier página web, aparece un aviso solicitando la aceptación de *cookies*. Sin embargo, la mayoría de los usuarios acepta sin leer ni comprender lo que está validando, motivados por el deseo inmediato de acceder al contenido.

Fedorov sostiene que con la inteligencia artificial ocurrirá lo mismo que con las cookies: se notificará que se está interactuando con un sistema de IA, pero no se analizará ni se cuestionará la veracidad de los resultados generados.

Según el autor, este enfoque superficial omite una cuestión fundamental: detrás de cada sistema hay personas —o grupos de personas— que deciden qué datos introducir, qué criterios aplicar y cómo utilizar los resultados. Publican el contenido generado y, en muchos casos, se benefician directamente de él.

Por este motivo, Fedorov propone que se introduzca una figura clara de "autor humano responsable", una persona identificable que asuma la responsabilidad de las decisiones asociadas al desarrollo, implementación y uso del sistema de IA. Esta medida, además de fortalecer la rendición de cuentas, podría reforzar la confianza social y facilitar el control ético y legal sobre los sistemas automatizados.

Un ejemplo reciente que ilustra esta preocupación es la huelga del sindicato de guionistas de Hollywood (Writers Guild of America) en 2023. Uno de los motivos centrales del conflicto fue el temor a que sistemas de inteligencia artificial generativa, como ChatGPT, pudieran sustituir a guionistas y escritores profesionales. Tras meses de negociación, se alcanzó un acuerdo con los estudios cinematográficos en el que se reconoce que los escritores pueden utilizar herramientas de IA como apoyo en su trabajo, pero el crédito de autoría y la contratación siguen recayendo en el ser humano. Es decir, la IA puede ser una herramienta, pero no sustituye al autor humano.

Según Vadim Fedorov, este modelo es un ejemplo acertado de cómo debe abordarse el problema: no se trata de reemplazar al ser humano con una máquina, sino de permitir que el autor humano mejore su trabajo mediante el uso de herramientas basadas en inteligencia artificial. Esta perspectiva preserva la responsabilidad, la creatividad y la trazabilidad de la autoría, al tiempo que incorpora las ventajas tecnológicas del desarrollo automatizado.

En este caso, la cuestión no gira en torno a la preservación de los derechos de autor sobre los contenidos que las compañías utilizan para entrenar sus sistemas de inteligencia artificial, sino que se centra en una dimensión distinta pero igualmente crucial: la necesidad de que exista una persona responsable de los resultados que genera el sistema.

Tal como advierte Vadim Fedorov, debe haber un responsable humano que asuma la autoría o supervisión de las salidas generadas por herramientas como *ChatGPT* u otros modelos generativos. Esto se vuelve especialmente relevante si se considera que los resultados que ofrecen estos sistemas pueden situarse, en ocasiones, más cerca de la desinformación que del conocimiento preciso y contrastado.

Por ello, la figura del "autor humano responsable" adquiere un papel clave: no se trata solo de supervisar técnicamente la IA, sino de asumir la responsabilidad ética, legal y social sobre el uso y las consecuencias del contenido generado.

Por último —y aunque pueda parecer una cuestión ya resuelta, en realidad no lo está—, las compañías que desarrollan y comercializan sistemas de inteligencia artificial deberían ofrecer explicaciones mucho más claras y accesibles sobre la finalidad concreta de dichos sistemas. Arvind Narayanan, profesor de informática de la Universidad de Princeton, en una entrevista a un medio de comunicación sostuvo “the main thing that was wrong with generative AI last year is that companies were releasing these really powerful models without a concrete way for people to make use them” (El Mundo, 2024).

Es más, el profesor Arvind Narayanan, junto a Sayash Kapoor, en un libro publicado en 2024, advierte que el foco del debate debería desplazarse: en lugar de centrarnos exclusivamente en lo que la inteligencia artificial puede hacer por sí sola, deberíamos estar mucho más preocupados por lo que las personas harán con ella.

En este contexto, la transparencia debería extenderse también a las personas que utilizan los sistemas de inteligencia artificial en cualquiera de las etapas de su ciclo de vida. Esto permitiría que, en caso de que se produzca un riesgo o daño, pueda identificarse a la persona responsable de su manejo o implementación.

No obstante, esta exigencia debe conciliarse con la necesaria protección de la privacidad de los usuarios y operadores. Es decir, se trataría de establecer mecanismos de trazabilidad y rendición de cuentas que no vulneren los derechos fundamentales, sino que garanticen un equilibrio entre responsabilidad individual y protección de datos personales.

Como puede observarse, el debate en torno a la transparencia en los sistemas de inteligencia artificial está plenamente abierto. Aún es necesario recorrer un trecho importante para identificar con precisión las lagunas e incertidumbres jurídicas que la rodean, así como para determinar cuáles son las soluciones más adecuadas y realistas desde el punto de vista técnico, legal y ético.

Pero, por encima de todo, urge clarificar qué se espera realmente del principio de transparencia: si ha de ser un mero requisito formal o, por el contrario, una

herramienta efectiva para salvaguardar a las personas frente a los riesgos que plantea la inteligencia artificial.

6. REFERENCIAS

Adobe (2025, 24 de abril). *Directrices para usuarios de la IA generativa de Adobe*. <https://www.adobe.com/es/legal/licenses-terms/adobe-gen-ai-user-guidelines.html>

Agencia Española de Protección de Datos (2023, 20 septiembre). *Inteligencia artificial: Transparencia*. <https://www.aepd.es/prensa-y-comunicacion/blog/inteligencia-artificial-transparencia>

Blasi, Cristina y Cañabate, Josep (2024). *Legislación y derecho digital para no juristas*. UAB Servei de Publicacions. <https://publicacions.uab.cat/llobres/legislacion-derecho-digital-para-no-juristas>

Colomina, Carme y Galceran-Vercher, Marta (2024). “Las otras geopolíticas de la inteligencia artificial”. *Revista Cidob d’Afers Internacionals*, nº 138. <https://www.cidob.org/publicaciones/las-otras-geopoliticas-de-la-inteligencia-artificial>

Comisión Europea (2019). *Directrices éticas para una IA fiable*. Oficina de Publicaciones de la UE. <https://op.europa.eu/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>

COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES Estado de la Década Digital 2024. COM/2024/260 final. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52024DC0260>

Coulter, Martin (2024, 13 de junio). EU’s new AI rules ignite battle over data transparency. *Reuters* <https://www.reuters.com/technology/artificial-intelligence/eus-new-ai-rules-ignite-battle-over-data-transparency-2024-06-13/>

Council of Europe. *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*. Vilnius, 5.IX.2024. <https://rm.coe.int/1680afae3c>

Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital 2023/C 23/01. DO C 23 de 23.1.2023. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=OJ:JOC_2023_023_R_0001

Diario La Ley (2024, 8 de marzo). *Entrevista a Alberto Pena Fernández*. <https://diariolaley.laleynext.es/dll/2024/03/08/el-ecat-es-el-brazo-cientifico-que-da-soporte-al-legislador-europeo-para-supervisar-el-funcionamiento-de-los-algoritmos-utilizados-por-las-plataformas-digitales>

El Mundo (2024, 30 de diciembre). *In 2024, artificial intelligence was all about putting AI tools to work*. <https://www.mundoamerica.com/news/2024/12/30/6772a1e6e85eced70f8b457a.html>

El Tiempo (2025, 29 de enero). *Italia pide explicaciones a DeepSeek sobre la utilización de datos personales en su nueva IA*. <https://www.eltiempo.com/mundo/europa/italia-pide-explicaciones-a-deepseek-sobre-la-utilizacion-de-datos-personales-en-su-nueva-ia-3421903>

Ep Data (2024, 10 de abril). *El uso de la inteligencia artificial en las empresas de la Unión Europea, en cifras*. https://www.epdata.es/datos/inteligencia-artificial-espana-union-europea-empresas/727?utm_source=chatgpt.com

Ericsson (2024, 19 de septiembre). *Open letter: Europe needs regulatory certainty on AI*. <https://www.ericsson.com/en/news/2024/9/open-letter-on-fragmented-regulation-risks-to-eu-in-ai-era>

European Court of Auditors (2024). *EU Artificial intelligence ambition Stronger governance and increased, more focused investment essential going forward*. Special Report 08. https://www.eca.europa.eu/ECAPublications/SR-2024-08/SR-2024-08_EN.pdf

Federal Office for Information Security (2024). *Transparency of AI Systems*. White paper. https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Whitepaper_Transparency_AI-Systems.pdf?__blob=publicationFile&v=2

Fedorov, Vadim (2024, 14 de marzo). "Responsible human author". <https://vf42.com/responsible-human-author/>

Fernández, Carlos y Eguiluz, Josu Andoni (2024, 28 de junio). “Diez puntos críticos del Reglamento europeo de Inteligencia Artificial”, Diario LA LEY, n. 85. https://diariolaley.laleynext.es/Content/Documento.aspx?params=H4sIAAAIAAAEAC2Ny6rCQAxAv8YspVrxchfZ1G4EEdFy9-k0tIFxRmfSav-xkfgkAcnyX3kNDf8VLyKF6UEeQ4xzFds0sig1GYsN4sfVxgrYw3kdCRfR4flq5aJG2pNgpg6TtWMBWhU8mfOuNpufyEP8XGkSXpSiaGi9LktXYd1U1hsStNKMdhLEBPeg7KMEg_HAz9-JkpueFEPeM-iJO4pHx7fsfVqGqrrYbLuwfnLdekVCPPofs-Qc2AwcT7wAAAA=WKE

Galcerán, Marta (2024). *El Pacto digital global y la transición responsable e inclusiva*. Cidob Report #12, 2024. <https://www.cidob.org/publicaciones/el-pacto-digital-global-y-la-transicion-responsable-e-inclusiva>

Gavilán, Ignacio G.R. (2022, 7 de febrero). Cuatro principios para una buena explicabilidad de los algoritmos. *Blue chip*. <http://bluechip.ignaciogavilan.com/2022/02/cuatro-principios-para-una-buena.html>

Gavilán, Ignacio G.R. (2024, 17 de enero). La explicabilidad en los grandes modelos de lenguaje. *Blue chip*. <http://bluechip.ignaciogavilan.com/2024/01/la-explicabilidad-en-los-grandes.html>

Infobae (2025, 29 de enero). *DeepSeek desaparece de Italia por preocupaciones por la privacidad de los datos*. <https://www.infobae.com/tecno/2025/01/29/deepseek-desaparece-de-italia-por-preocupaciones-por-la-privacidad-de-los-datos/>

IT Reseller (2024, 20 de junio). *Dos de cada tres españoles ya están familiarizados con la inteligencia artificial*. <https://www.itreseller.es/en-cifras/2024/06/dos-de-cada-tres-espanoles-ya-estan-familiarizados-con-la-inteligencia-artificial>

Jarowsky, Luiza (2024, 24 de septiembre). “AI regulation is not uncertain”. *Luiza’s Newsletter*. <https://www.luizasnewsletter.com/p/ai-regulation-is-not-uncertain>

Le Maire, Bruno (2024, 15 de febrero). “Déclaration sur l’intelligence artificielle”. *Vie publique (République Française)*. <https://www.vie-publique.fr/discours/293068-bruno-le-maire-15022024-intelligence-artificielle>

Naciones Unidas. Asamblea General. *Aprovechar las oportunidades de sistemas seguros, protegidos y fiables de inteligencia artificial para el desarrollo sostenible*, A/78/L.49. 11 de marzo de 2024.

https://agatadata.com/wp-content/uploads/2024/11/Aprovechar-las-oportunidades-de-sistemas-seguros-protegidos-y-fiables-de-inteligencia-artificial-para-el-desarrollo-sostenible-11_03_24-Naciones-Unidas.pdf

Naciones Unidas. Asamblea General. *Resolución aprobada por la Asamblea General el 22 de septiembre de 2024 [sin remisión previa a una Comisión Principal (A/79/L.2)] 79/1. El Pacto para el Futuro*. <https://docs.un.org/es/A/RES/79/1>

Narayanan, A. & Kapoor, S. (2024). *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell The Difference*. Princenton University Press.

<https://press.princeton.edu/books/hardcover/9780691249131/ai-snake-oil?srsId=AfmBOop0dJ1qQ2I3x5WT6VyPu0d5STsLMhkoxfycz8JfvO43KMrBTOWO>

Navas, Susana (2023). *ChatGPT y modelos fundacionales. Aspectos jurídicos de presente y de futuro*. Reus editorial. <https://www.editorialreus.es/libros/chatgpt-y-modelos-fundacionales-aspectos-juridicos-de-presente-y-de-futuro/9788429027907/>

Pathak, Amrita (2024, 14 de mayo). "¿Qué son los sesgos algorítmicos y cómo detectarlos?" *Geekflare.com*. <https://geekflare.com/es/algorithmic-biases/>

Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (...) (Reglamento de Inteligencia Artificial).

<http://data.europa.eu/eli/reg/2024/1689/oj>