PROPUESTA DE BASES PARA LA AUTORREGULACIÓN DEL USO DE INTELIGENCIA ARTIFICIAL EN LA CONTRATACIÓN PÚBLICA

Adán Nieto Martín

Instituto de Derecho penal Europeo e Internacional

Administración Pública e Inteligencia Artificial: regulación y uso de la AI en el ámbito de la contratación pública" (TED2021-130682B-I00), financiado por MCIN/AEI/10.13039/501100011033 y por la Unión Europea NextGenerationEU/PRTR

Este documento tiene un marcado carácter ejecutivo, lo que explica el lenguaje utilizado y su carácter esquemático. Su finalidad es sintetizar los resultados obtenidos en el Proyecto de Investigación APIA y busca servir de orientación a los operadores jurídicos en su función de autorregular el uso de IA en el ámbito de la contratación pública

1.- PRESUPUESTOS

La propuesta de modelo de autorregulación que a continuación va realizarse descansa en cuatro elementos. Primero, las exigencias procedentes del Reglamento 2024/1681 para los productores de alto riesgo. Se trata del estándar general, pues el Reglamento incita a los productores de sistemas de bajo riesgo que de manera voluntaria se adapten a las mismas. Segundo, los estándares más importantes en materia de IA. La autorregulación regulada, que es la estrategia regulatoria que sigue el Reglamento, combina la capacidad de autorregulación de sus destinatarios, con los estándares que van generando diversas instituciones y agentes, de ahí que sea imprescindible tenerlos presentes. Tercero, en lo que concierne a los destinatarios, nos centraremos en las obligaciones de los productores o proveedores, en el confuso lenguaje que adopta el Reglamento. Cuarto, la metodología y estructura que se sigue en la elaboración del modelo está basada en los programas de cumplimiento normativo.

La regulación de la IA por parte del Reglamento 2024/1689 descansa, como es bien conocido, en la siguiente tripartición: sistemas de IA prohibidos, sistemas de alto riesgo y de bajo riesgo. El núcleo de la regulación se concentra en los segundos. En relación, a los primeros se prohíbe sin más su producción (art. 5), en lo que concierne a los de bajo riesgo se confía más en la autorregulación voluntaria; además de cumplir una serie de obligaciones básicas (transparencia, etiquetado, prohibición de engaño) que recuerdan a la regulación de los productos "ordinarios", se les alienta a que adopten voluntariamente códigos de conducta. Además, cuando un modelo de bajo riesgo se utiliza como base para un sistema de alto riesgo, debe cumplir con idénticos requisitos, especialmente el contar con una organización post venta orientada a la mitigación del riesgo.

Para comprender la estructura de la regulación del Reglamento IA debe igualmente tenerse presente que, aunque a lo largo de este trabajo vamos a centrarnos en las obligaciones que establece para los productores (proveedores), el Reglamento IA contiene también obligaciones para otros agentes económicos que intervienen en la cadena de valor de un producto IA, como son los importadores o distribuidores. Lo novedoso es que contiene también obligaciones de prevención muy importantes para los denominados "responsables de despliegue", es decir, aquellas organizaciones públicas o privadas que utilizan sistemas IA en sus actividades profesionales.

El posible que en una misma organización puedan darse ambas características. Sería el caso por ejemplo de una Universidad que provee, es decir, produce sistemas de IA, pero que a su vez utiliza sistemas de inteligencia artificial que ha adquirido a otros proveedores en distintas parcelas de su actividad. Esta confluencia de roles conlleva una complejidad regulatoria notable.

El Reglamento 2024/1689 utiliza la fuerza de autorregulación de sus diversos destinatarios, imponiéndoles un marco de obligaciones muy general y permitiendo después que cada obligado lo adapte a su organización y al nivel de riesgos que desarrolla. Es un sistema bien conocido y que el legislador utiliza en áreas tan dispares como la protección de riesgos laborales, el blanqueo de capitales o la protección de datos. Esta

forma de regulación, a diferencia de otras tradicionales, integra, más bien necesita, normas de *soft law* y estándares, cuya función es orientar y ayudar a los destinatarios de las obligaciones de autorregulación. Por esta razón no puede prescindirse de los *Ethical Frameworks* más importantes existentes a la hora de concretar el modelo de autorregulación que se propone. El Reglamento de IA es consciente de la necesidad de contar con normas de *soft law* y estándares dedica el Capitulo X a la elaboración de este tipo de normas, que son más importantes en los sistemas de bajo riesgo, basados en la colaboración público-privada. Dentro de los estándares existentes, destacan las Directrices éticas para una IA fiable, elaboradas en 2019 por un grupo de expertos creado por la Comisión Europea.

La propuesta de modelo de organización que a continuación va a realizarse estructura los diversos requisitos que establece el Reglamento 2024/1689 a partir la metodología que proporcionan los modelos de cumplimiento normativo. Los modelos de cumplimiento normativo y sobre todo los que se han desarrollado en Derecho penal, a partir de la responsabilidad penal de las personas jurídicas, tienen el mérito de que han sabido condensar bajo una estructura unitaria diversas estrategias de autorregulación que históricamente se han sucedido de manera independiente, lo que permite estructurar, pero también rellenar los diversos contenidos necesarios para que funcionen los sistemas de autorregulación que establece el Reglamento IA. Como en cualquier otro espacio de la autorregulación regulada, el Reglamento IA impone un conjunto de requisitos, pero no proporciona un marco para estructurarlos. Este punto lo deja en manos de los destinatarios de la norma.

La conexión con el cumplimiento normativo aporta, en síntesis, tres ventajas.

La primera metodológica. Tal como se está indicando, complementa y desarrolla los espacios que debe tener todo modelo de organización en IA y estructura sus diversos elementos. De manera resumida puede decirse que un programa de cumplimiento normativo cuenta con cuatro grandes bloques, en que se encuadran sus diversos componentes. El primero es un elemento inmaterial, conformado por la cultura de la organización. Sin un mínimo de valores comunes, de ética organizacional, ningún programa de cumplimiento funciona, salvo que la organización se convierta en un panóptico. El segundo está compuesto por elementos preventivos, que tienen como cometido identificar riesgos, establecer normas de conducta y procedimientos de control. El tercero es la parte reactiva. Es aquí donde el modelo de cumplimiento normativo IA presenta una mayor originalidad, tal como ya se ha indicado. La cuarta es la parte institucional u organizativa. El Reglamento IA pone especial énfasis, como suele ser normal en las normas jurídicas que utilizan la autorregulación, en las obligaciones documentales, pero ha descuidado un tanto la parte relativa a la asignación de responsabilidades dentro de la organización.

De aquí deriva, precisamente, la segunda ventaja de poner en conexión el sistema autorregulatorio IA con el cumplimiento normativo: ello puede propiciar encuadrar estas nuevas obligaciones dentro de la estructura organizativa de la entidad. La propuesta que aquí se haría es ubicar la gestión del sistema de autorregulación IA dentro de los departamentos de cumplimiento normativo. En grandes y medianas empresas ha

aparecido ya una organización que aúna las exigencias de autorregulación que pueden proceder de diversos sectores del ordenamiento o de compromisos adquiridos voluntariamente por la entidad. La autorregulación IA podía ser de este modo un apéndice más de los departamentos de cumplimiento, lógicamente con sus propios responsables y dueños del control, pero encajada organizativamente en este sector de la organización empresarial o de otro tipo.

En realidad, y esta es la tercera de las ventajas, la utilización del cumplimiento normativo como marco conceptual y metodológico para responder a las obligaciones del Reglamente es una forma de cumplir con lo dispuesto en su art. 17, el establecimiento de un sistema de gestión de calidad. Este sistema de gestión de calidad coincide a grandes rasgos con la idea de cumplimiento normativo. Pues lo que pretende es ordenar el conjunto de medidas de autorregulación que impone el reglamento e incardinarlas dentro del sistema de gestión de la organización del productor (proveedor). Por esta razón, incorpora (art. 17 a) la elaboración de una estrategia para el cumplimiento de la norma, obliga a establecer procedimientos que acompañen todo el proceso de elaboración (art. 17.1 d), un sistema de responsabilidades interno (art. 17 1 m), la evaluación del riesgo (art. 17. 1 g). Como a continuación va a mostrarse, la proyección del cumplimiento normativo sobre esta obligación le proporciona una mayor claridad, conectando sus diversos componentes de una forma más sencilla y coherente.

Las particularidades que tiene la autorregulación en el caso del modelo IA no empecen a adoptar su construcción a partir de la metodología del cumplimiento normativo. El que la autorregulación IA deba centrarse tanto en la prevención de daños, como en la supervisión y corrección de daños ya acaecidos no representa ninguna particularidad. Pues se trata esencialmente de arbitrar procedimientos y establecer responsabilidades de control en las personas naturales que componen la organización que atiendan a ambos cometidos. Los programas de cumplimiento cuentan además con una parte reactiva que es muy importante. El cumplimiento normativo no solo tiene como función la prevención, por ejemplo, de la corrupción, sino también si esta se produce, tiene que articular mecanismos de detección y sanción. Por esta razón son parte de los programas de cumplimiento elementos predominantemente reactivos como los canales de alertas, la realización de investigaciones internas o la imposición de sanciones disciplinarias. A ellos habrá que sumarles ahora sistemas destinados a la detección y mitigación. La IA no es el único caso. La aparición de la diligencia debida en sostenibilidad, obra de la Directiva 2024/1760, también incide en un modelo de cumplimiento normativo donde la detección y la mitigación predominan sobre los aspectos preventivos.

II.- ELEMENTOS

Aclarados los fundamentos sobre los que se estructura nuestra propuesta, un esquema general del cumplimiento normativo en materia de IA destinado a un productor/responsable de despliegue sería el siguiente:

A. Valores y principios de la organización

B. Elementos preventivos.

1. Código de conducta
2. Análisis de riesgos (evaluaciones de impacto algorítimico)
3. Sensibilización y formación de empleados
4. Principios inspiradores de los procedimientos de control
    i. Trazabilidad y explicabilidad del sistema con el fin de conocer sus decisiones.
    ii. Supervisión
    iii. Calidad de los datos y entrenamiento.
5. Medidas de ciberseguridad

C. Elementos de supervisión.

1. Evaluación continua de la calidad del sistema y de sus datos.
2. Flujos de comunicación con otros miembros de la cadena de valor y con los usuarios
3. Planes de retroceso (fallback)
4. Medidas de reparación.

D. Elementos institucionales y organizativos.

1. Responsabilidad del órgano de dirección de la organización.
2. Responsabilidad del IA encargado de IA.
3. Trazabiliadd y documentación

A. *Valores y principios: la cultura de la organización*.

En materia de valores y principios en materia de IA y cultura de la organización debe distinguirse entre dos planos. El primero es el de los valores éticos de la organización en sí. El segundo es el de los valores éticos que deben marcar el desarrollo y la producción de la IA. El Reglamento de IA y los marco éticos se refieren a estos últimos, pero dejan de lado el marco ético general de la organización en que se insertan. Este último resulta imprescindible. La argamasa de todos los sistemas de cumplimiento es la cultura de la organización. Es un elemento inmaterial y por tanto complejo de introducir y mucho más de valorar, medir, en el mundo jurídico, pero resulta imprescindible tanto para que las normas, en este caso la regulación de la IA, sean realmente efectivas, como para que la producción y el desarrollo de la IA se adecúe a los principios éticos que figuran en los diversos Ethical Frameworks. Un IA legal y ética, como por ejemplo proponen las Directrices Éticas para una IA fiable, no tiene sentido si no se enmarca en una organización ética. La primera tarea por tanto que debe acometer un proveedor es alinear los valores que se fijan en la regulación y estándares éticos con los de su organización. Como señalan las Directrices éticas "para garantizar la fiabilidad de la IA es necesario, más allá de desarrollar un conjunto de normas, crear y mantener una cultura y una mentalidad ética a través del debate público, la educación y el aprendizaje práctico".

Generar una determinada cultura de la organización no es fácil. Pero el cumplimiento normativo nos ha enseñado, la importancia que en este punto tiene el liderazgo; la participación de los destinatarios en la confección de los distintos protocolos y normas

de conducta que componen el sistema de autorregulación, la denominada justicia procedimental es un componente básico del cumplimiento normativo; la existencia de sanciones y recompensas; la investigación de las infracciones; la formación etc..

B. *Elementos preventivos*.

a. Códigos éticos.

La convergencia entre los valores de la organización y los valores IA debe concretarse, visualizarse, en la confección del código ético de la entidad. Esta coherencia o continuidad puede materializarse a efectos prácticos del siguiente modo. De un lado, los códigos éticos deberán abrir un nuevo capítulo destinado a la IA. La estructura de los códigos éticos es normalmente la siguiente: (1) descripción de valores generales de la organización, (2) asignación de responsabilidades en la ejecución del código ético; (3) concreción de estos valores en una serie de espacios: protección de los trabajadores, del medio ambiente, competencia, relaciones con los funcionarios públicos, proveedores… Pues bien, en esta suerte de "parte especial" del código habrá de añadirse un capítulo nuevo dedicado a recoger de manera sucinta los principios básicos que inspiran la producción de IA. Después, como ocurre en tantas otras materias del cumplimiento normativo, estos principios pueden ser desarrollados en un documento, en una política específica sobre esta materia.

Siendo más precisos, esta parte especial del código ético dedicada a la IA, deberá contar con dos tipos de disposiciones. Una primera, que afectará a prácticamente todos las organizaciones tendrá que establecer las normas de conducta relativas a la utilización de sistemas de inteligencia artificial por parte de todos sus miembros. Obviamente estas normas serán muy diversas, no es lo mismo un hospital, que una universidad que un despacho de abogados. La segunda parte, que es ahora la que más interesa, afecta a los productores (proveedores) de IA y tiene que ver con las características de sus productos. Son estas últimas normas éticas las que se contienen en los diversos marcos éticos. Ambos aspectos pueden desarrollarse después en políticas. Por ejemplo, tendría todo el sentido de que las normas relativas al uso de la inteligencia artificial dentro de una organización se concretaran en una política similar a las que actualmente existen relativas a los TIC o incluso se incluyeran dentro de estas, en cuanto que existe una clara continuidad normativa entre unas y otras.

La incorporación del código ético IA al código ético general en el marco del cumplimiento normativo tiene como valor añadido que juridifica sus disposiciones. En materia de cumplimiento normativo se ha aclarado que los códigos éticos equivalen a ordenes de trabajo que el empresario da a sus empleados, en virtud de su capacidad de dirección y organización que nace del contrato de trabajo. Ello implica la posibilidad de aplicar las sanciones disciplinarias laborales al incumplimiento de los deberes éticos y las normas de conducta que de ellos se deriven en materia de IA.

En cualquier caso, los diversos valores incorporados al código ético, como en seguida va a comprobarse, resultan relevantes para realizar el análisis de riesgos y, a partir de aquí, conformar los diversos controles que han de insertarse en los procedimientos de producción.

b. Análisis de riesgos (o evaluación de impacto algorítmico).

El análisis de riesgos es un elemento medular e indispensable de cualquier programa de cumplimiento y representa una obligación legal que se desprenden de todas las normas legales que utilizan la autorregulación regulada. La metodología es común e incluso existe un norma IS0 que establece un modelo de gestión unitario para la realización del análisis de riesgos. El art. 9 del Reglamento es el que hace referencia a este aspecto. Su contenido, algo caótico y confuso, se entiende mejor cuando se lee a partir de la estructura tradicional del análisis de riesgos: (1) fijar las obligaciones legales y éticas a las que se refieren los riesgos de incumplimiento (2) establecer la probabilidad del riesgo en abstracto de un determinado sistema IA, (3) determinar de los factores que pueden incrementar su aparición, (4) examinar la idoneidad de los controles ya existentes, (5) proponer en su caso nuevos controles que reduzcan el riesgo.

El primer paso, la base del análisis de riesgos, requiere concretar los riesgos y los daños que se quieren evitar. El art. 9.2 señala, en relación con los proveedores/productores como riesgos la salud, la seguridad o los derechos fundamentales. Para determinar estos conceptos resultan de gran utilidad las indicaciones y el marco que proponen los diversos estándares éticos, pero sobre todo el marco de referencia debiera ser el código ético de cada organización donde estos valores se concreten y la política IA, en caso de que se haya considerado oportuno desarrollarla. Sintéticamente podríamos señalar que el análisis de riesgos debe tener como objetivo el evitar el daño a la vida, la salud y otros derechos fundamentales básicos, lo que se produce sobre todo cuando existen sesgos o cuando no respeta la autonomía de los usuarios, manipulándolos.

El Reglamento IA en su art. 6 y en los Anexos I y III al definir los sistemas de IA de alto riesgo nos proporciona una gran orientación, que puede ser útil para concretar el análisis de riesgos. El Anexo I se refiere sobre todo al uso de la IA en productos industriales muy variados que han sido regulados por el legislador europeo (juguetes, medios de transporte productos sanitarios..). El riesgo prototípico en estos casos será la vida y la salud de las personas. El Anexo III se refiere a la realización de determinadas actividades a través de la IA (sistemas de identificación biométrica, destinados a ser utilizados al acceso o la admisión de personas físicas a centros educativos o evaluar resultados; gestión de contratos de trabajo etc..) donde el riesgo es para otros derechos fundamentales, y especialmente la no existencia de sesgos que produzcan discriminaciones. Un riesgo transversal es que la IA no respete la autonomía del usuario y lo manipule. En la evaluación de riesgos es útil igualmente atender a las prácticas prohibidas por la IA (art. 5). Obviamente la creación intencional de una IA, por ejemplo, para explotar las vulnerabilidades de una personas física o colectivo está prohibida, pero la prohibición de determinadas IA tiene un carácter objetivo, y además debe prevenirse, como veremos, el que un sistema IA generado para otras finalidades, sea utilizado por los usuarios para alguna de las finalidades prohibidas. Por ello, y de manera coherente, el art. 99 del Reglamento al establecer las sanciones obliga a sancionar tanto la creación de IA prohibida dolosa, como imprudente. La intencionalidad o negligencia sólo son factores de graduación de la sanción (art. 99. 7 g).

El art. 9 del Reglamento aun sin decirlo parte de que los controles deben ser proporcionales al riesgo. No se exige un riesgo cero, sino un control razonable, como se desprende de alguna de las expresiones que emplea como "mitigarse o eliminarse razonablemente", "riesgos residuales". Ello implica que la implementación de controles debe hacerse sobre la base de un análisis de coste y beneficio y valoración del riesgo (art. 9.3). Análisis de riesgos y valoración del riesgo son conceptos diferentes. La valoración del riesgo consiste en priorizar unos riesgos sobre otros atendiendo a su previsibilidad y al grado de impacto que puedan tener en la salud, la seguridad o los derechos fundamentales (art. 9.4).

El aspecto más innovador del análisis de riesgos en esta materia deriva de la característica principal del "producto IA": la autonomía e imprevisibilidad ex ante de los out puts del sistema. Todo análisis de riesgo debe revisarse cuando existe algún tipo de incumplimiento que revela la inadecuación de los controles existentes, más aquí esta actividad es esencial, es cualitativamente distinta. De este modo, el análisis de riesgos es una actividad no ya periódica, como ocurre en el resto de sectores, sino constante, ininterrumpida, de un proceso "continuo y planificado durante todo el ciclo de vida de un sistema" (art. 9.2). Y para ello se exige una alta coordinación entre los encargados de la supervisión del producto IA (el sistema de vigilancia post comercilización) y los encargados del análisis de riesgo.

Sin duda en este punto, el aspecto más complicado y polémico, es la necesidad de incluir en el análisis de riesgos y por tanto, la necesidad de establecer medidas de mitigación, relativas a los usos ilícitos que puedan ser razonablemente previsibles (art. 9.2 d). El control de un mal uso por parte de los usuarios debe limitarse con medidas de gestión de riesgos que deben adoptarse tanto en la fase de diseño, como durante medidas posteriores como singularmente la formación de los usuarios. Especialmente este análisis de riesgos será complicado en los denominados "modelos de IA de uso general" diseñados para atender a una gran variedad de tareas. En este punto la concreción de los riesgos ex ante resulta muy compleja, por lo que aquí tendrá especialmente importancia la revisión constante de los riesgos del sistema a la luz de las informaciones que vayan aportando los controles de supervisión post venta.

A la vista de cuanto acaba de indicarse el análisis de riesgos IA tiene un fuerte componente reactivo. Normalmente, en el análisis de riesgos se intentan anticipar los riesgos, mientras que en el caso de la producción de productos IA, ante la imprevisibilidad de su comportamiento y ante la necesidad de corregir comportamientos de terceros, en análisis de riesgos requiere estar atentos a diversas informaciones que permiten concretar y reducir la incidencia de un riesgo a través de la modificación del sistema de IA.

c. Formación y sensibilización.

Los deberes de formación, la alfabetización en IA son un deber general que se impone tanto a los proveedores como a los responsables del despliegue (art. 4). La formación en cumplimiento normativo es siempre adecuada y proporcional al riesgo, aspecto este que no se menciona expresamente en el art. 4. No tiene sentido una formación uniforme, debiéndose acomodarse a las características del usuario; es decir, y pensando en los

responsables del despliegue a las características del puesto de trabajo de los empleados y al nivel de riesgo asociado al desempeño de sus funciones. Distinta a la formación es la sensibilización, cuyo objetivo es la compresión de los valores y promover la reflexión ética. Este aspecto no se menciona en el art. 4 que parece más orientado a una formación técnica, sin embargo, como ya se indicó en relación con la cultura de la organización sin fomentar la reflexión ética sobre los problemas de la IA no existirá un marco adecuado para que los controles previstos en el Reglamento pueda funcionar y que este cumpla con sus objetivos.

d. Procedimientos.

El desarrollo de un sistema de IA y de los distintos procedimientos que lo componen debe tener como finalidad, como ocurre con cualquier otro producto, la no producción de daños, especialmente si estos pueden afectar a derechos fundamentales, como ocurre en los casos en que el sistema contiene sesgos, y el respeto a la autonomía de los usuarios. Escapa totalmente a mis capacidades indicar cuáles son los procedimientos más idóneos en la fabricación de IA tendentes a neutralizar estos riesgos, pero si resulta necesario señalar al menos tres criterios transversales que deben ser de aplicación: la explicabildiad, la posibilidad de supervisión y la calidad de los datos.

La explicabilidad técnica de un sistema implica que las decisiones que adopte puedan ser comprensibles para sus usuarios y que estos tengan capacidad para rastrearlas. La exigencia de explicabilidad aumenta en función de los riesgos del sistema: a mayor capacidad de impacto en los derechos de las personas mayor debiera ser la explicabilidad. Si los *out puts* del sistema funcionan como una caja negra su previsibilidad y corrección será mucho más compleja. Por esta razón, un sistema de IA basado en el tratamiento de datos a través de una metodología puramente estadística es mucho menos conforme con este objetivo (sistemas estadísticos), que los sistemas basados en la ingeniería del conocimiento (sistemas cognoscietivos), donde los criterios que utiliza el sistema son muchos más rasteables y sobre todo pueden ser explicados a través de un tipo de razonamiento que es similar al razonamiento humano. Los sistemas cognoscitivos de IA se construyen sistematizando y articulando el conocimiento que existe en una materia en una taxonomía, con la finalidad de que el sistema imite este tipo de razonamiento.

El segundo elemento transversal es la supervisión o control humano. Desde luego, no tendría sentido alguno que se exigiera que un humano estuviera siempre al control con el fin de evitar daños. Esto acabaría con la utilidad social de la IA. Pero al menos debe exigirse una posibilidad de intervención o supervisión en el sistema. La independencia no puede ser total o excluyente de la intervención humana. Siempre debe ser posible que el usuario en un momento determinado puede decidir si quiere utilizar el sistema o suspenderlo, a través de un "botón de desconexión" que permita tomar al humano totalmente el control sobre las operaciones. Nuevamente los controles que faciliten o incluso requieran la supervisión deben ser proporcionales a la probabilidad y la intensidad del daño (art. 14.3). El art. 14 del Reglamento 2024/1689 concreta esta obligación en su art. 14 indicando que "los sistemas IA de alto riesgo se deseñarán y desarrollarán de modo que puedan ser vigilados de manera efectiva por personas físicas durante el periodo que estén en uso", con el fin de "reducir al mínimo los riesgos para la

salud, la seguridad o los derechos fundamentales". La supervisión, aunque debe ser facilitada por el proveedor/productor, correrá a cargo normalmente del usuario/responsable del despliegue, ello obliga a que el sistema tenga un interfaz adecuado que permitan a sus usuarios intervenir adecuadamente (art. 14.4), de este modo, por ejemplo, un sistema de ayuda a la decisión (un sistema que recomiende al médico el tratamiento más adecuado para un paciente) debe incorporar *nudges* que eviten el sesgo de automaticidad, conforme al cual el usuario tiene a confiar plenamente en el sistema, pensando que todo está siempre en orden (art. 14.4 b).

El tercero tipo de procedimientos, recogido en el art. 10 del Reglamento, debe tener como finalidad garantizar la calidad de los datos y la fiabilidad del sistema, con el fin de eliminar los sesgos y la discriminación a la que pueden dar lugar unos datos inidóneos. Estos sesgos deben eliminarse en la fase de construcción del producto, para lo que es imprescindible además de controlar los datos un entrenamiento previo, que permita corregirlos. Los procedimientos que pueden establecerse para asegurar la calidad de los datos y que se transfieren correctamente al sistema (ingesta de datos) son muy variados una política interna en esta materia debiera establecer criterios acerca de las condiciones que deben tener los lugares de donde se recopilan los datos, los procedimientos que han de seguirse para su limpieza, su almacenamiento seguro; lo que conecta con las medidas de ciberseguridad que van a examinarse a continuación; el método que ha de seguirse para que su "ingesta" por el sistema sea correcto; y un sistema que permita evaluar los sesgos, lo que requiere realizar simulacros de impacto.

e. Medidas de ciberseguridad.

Los fabricantes de IA están obligados, finalmente, a establecer medidas de ciberseguridad, con el fin de prevenir ataques intencionados de terceros que quieran alterar su funcionamiento. No es una obligación que suela estar presente en otros productos. Los fabricantes de coches no tienen que diseñar un sistema de seguridad que impida que terceros distintos al conductor supriman, por ejemplo, sus sistemas de frenos. No obstante, la UE ha empezado a imponer obligaciones de ciberseguridad, a través de los denominados requisitos horizontales de ciberseguridad, a todos los productores de software y hardware por lo que no tendría sentido que no los impusiera a los de IA. Además los sistemas de inteligencia en ocasiones contienen datos personales sensibles, como datos biométricos, por lo que la obligación de protegerlos de ciberataques coincide con las que impone la normativa de protección de datos. En cualquier caso, la producción y la utilización de una IA poco robusta en ciberseguridad constituye un peligro para la salud, la seguridad y los derechos fundamentales, sobre todo cuando se trata de IA de carácter general que se utilizan en sectores donde un ataque puede dar lugar a un riesgo sistémico. En la fase de producción la alteración maliciosa de los datos con los que se está construyendo o entrenando el sistema supone un peligro que pertenece a la esfera de responsabilidad genuina de todo productor.

La cibercumplimiento es una parte de los programas de cumplimiento normativo con el que cuentan ya muchas empresas, mediante el que se protegen de amenazas de seguridad los datos de la empresa propios o los que debe proteger por la normativa de datos personales. Lo natural será que la ciberseguridad IA se integre en estos departamentos

C . *El sistema de vigilancia poscomercialización, planes de repliegue y reparación.*

Tal como advertíamos, en comparación con otros productos, el elemento más característico, genuino, de las medidas de precaución que tiene que implementar un productor/proveedor de IA es su supervisión tras su comercialización, con el fin de tener noticia cuanto antes de reacciones del sistema que pueden corresponderse a defectos de diseño o fabricación, pero también, aunque la IA haya sido construida con toda diligencia, debido a su actuación autónoma y la imprevisibilidad. Esta supervisión no debe confundirse con la supervisión humana a la que antes se hacía referencia, sino que se trata de una supervisión post-venta que debe establecer el productor.

Ello exige la implementación de toda una serie de procedimientos, a partir de lo dispuesto en el art. 72 del Reglamentos. Los describiré de manera secuencial o temporal, teniendo presente que los controles que se exponen deben ser siempre razonables y proporcionales al riesgo (art. 72.1).

El primero se desprende del modo en que se ha planteado la evaluación de riesgos: constante. Lo que implica que, aunque no exista ningún tipo de señal de peligro (*warning*), tras el análisis inicial, debe constantemente pensarse en los riesgos que puede generar el sistema. Este primer paso es particularmente importante, porque los sistemas de vigilancia poscomercialización deben ser proporcionales al riesgo. Lo que implica que en cuanto que este puede crecer o decrecer los controles que se establezcan deben ser dinámicos, alterables en función del riesgo. Parte del sistema de control de riesgos, puede ser un sistema de auditorias entre productores y usuarios profresionales (responsables del despliegue).

El segundo procedimiento de control poscomercialización es establecer un sistema de alerta temprana, que permitan al productor conocer en el menor tiempo posible cualquier indicio o resultado dañoso. Ello requiere que el sistema por sí solo o un supervisor autónomo generen los datos necesarios, para que pueda determinarse si existe un *out put* no querido. Desde luego diseñar el sistema para que facilite esta vigilancia ex post sería lo más adecuado, sobre todo cuando sus usuarios finales sean consumidores.

Por el contrario, cuando los usuarios sean profesionales, que lo utilizan para prestar sus servicios (responsables del despliegue), su cooperación en la recogida de estos datos resulta esencial. Ello deberá articularse principalmente a partir de las cláusulas contractuales entre el productor y el profesional en la que deben determinarse con la mayor precisión posible cuales son sus obligaciones legales. En estas clausulas contractuales habrá de determinarse como debe actuar el responsable del despliegue cuanto detecte alguna incidencia – evento lesivo-. Para ello habrá que crear protocolos de incidencias, a cuyo cumplimiento habrán de comprometerse contractualmente. A estos efectos puede ser útil la creación de un panel de control compartido que permita ver al proveedor en tiempo real las incidencias.

Un supuesto especial acaecerá cuando un sistema IA interactúe con otro en la ejecución de una tarea (art. 72.2). En este punto los dos productores deberán mapear en qué tomas de decisiones o actividades son independientes y fijar sus ámbitos de responsabilidad respectivos en estos casos a través de acuerdos de cooperación entre proveedores, en los

que bien pueden preverse, cuando no sea fácil deslindar ámbitos de responsabilidad, sistemas de monitorización integrada y protocolos de respuesta coordinada.

Desde luego, todo este entramado contractual será decisivo a la hora de establecer responsabilidades civiles y penales.

Bien puede ocurrir que pese a haber articulado correctamente un sistema de supervisión poscomercialización ocurra un evento dañoso, en este caso deben articularse procedimientos para la retirada del producto, plan de repliegue. A estos efectos lo primero que debe definirse con precisión es que se entiende por evento dañoso, término que bien puede coincidir con el de incidencia grave del art 73 del Reglamento IA y que viene definido en el art. 3 49): fallecimiento de personas, alteración de la gestión o del funcionamiento de estructuras críticas; afectación a los derechos fundamentales; daños al medio ambiente. No obstante, sería aconsejable que cada proveedor en su programa de cumplimiento determinara con mayor precisión que pude entenderse por incidencia grave a la vista de los riesgos que tiene cada sistema, utilizando también umbrales cuantitativos (¿cuántos incidentes deben existir para que consideremos que estamos ante una incidencia grave?). Los planes de repliegue exigirán nombrar expresamente en el Compliance IA quiénes son sus responsables, así como adoptar otras medidas de tipo operativo para bloquear parte del sistema, desconectándolo, o, en caso que no haya más remedio, establecer la retirada completa. En este punto también deben establecerse contractualmente cómo ha de articularse la cooperación con los responsables del despliegue.

Parte esencial del plan de repliegue es articular como se produce la comunicación con la autoridad de supervisión en caso de incidencia grave que se contiene en el art. 73 del Reglamento, para lo que existe un límite máximo de 15 días desde que se haya establecido o un vinculo causal o la probabilidad razonable de que esto ocurra, salvo que sea una infracción generalizada, caso en el que debe hacerse de inmediato. El margen y los criterios son tan amplios que convendrá establecer un procedimiento interno que concrete esta obligación. El Reglamento parece dar a entender que las acciones de retirada cuando se producen incidentes graves serán coordinadas, incluso dirigidas, por las autoridades de supervisión. Desde el punto de vista de la responsabilidad penal, la entrada de la autoridad administrativa implica en principio una desresponsabilización del fabricante si sigue sus instrucciones. Si bien en este punto debe tenerse en cuanta que en muchos casos los conocimientos del proveedor sean tan superiores a los de la autoridad administrativa que no debería poderse excusar en su toma de control para evitar daños que el proveedor, que es el responsable primario, podría haber evitado.

Finalmente debería reflexionarse se además de la “retirada” convendría también iniciar la reparación la mitigación de los daños. El Reglamento guarda silencio al respecto, pero algunos *Ethics Frameworks* como el de la UNESCO mencionan la necesidad de reparar el daño (puntos 54 ss.). Enfocado en los usuarios (consumidores) propone que los proveedores establezcan puntos de acceso para que pueda solicitarse rápidamente la reparación.

D. *Elementos institucionales y organizativos*.

a. Responsabilidades.

En el diseño de los programas de cumplimiento normativo, y por tanto en los IA Compliance programas, es esencial el reparto de responsabilidades. De acuerdo con la propuesta metodológica que inspira este trabajo partiremos en este punto de un sistema de institucionalización similar. Sus mimbres son los siguientes. El órgano de dirección de la entidad debe ser quien diseñe y apruebe las medidas de cumplimiento IA y las apruebe. Esta implicación se deriva de lo que indica el art. 31 bis 2 1º del Código penal (el órgano de administración ha adoptado y ejecutado con eficacia, antes de la comisión del delito, modelos de organización y gestión que incluyen las medidas de vigilancia y control idóneas) y se corresponde también con las disposiciones de la LSC, donde el consejo de administración es competente para la gestión de riesgos. La producción de IA representa un riesgo legal para una sociedad que tenga esta finalidad, por tanto, el consejo no puede desentenderse y debe abordar esta tarea. El Reglamento IA no contiene ninguna indicación a este respecto. Como cualquier otra norma que impone obligaciones a una sociedad estas se trasladan a su órgano de dirección, pero este podría delegarlas sin límites. El acoger el modelo "cumplimiento normativo" implica que el órgano de dirección asumirá la responsabilidad de diseño e implantación.

Esta obligación debe actualizarse. No se trata solo de intervenir en el momento de "puesta en marcha". De acuerdo, a las mejores prácticas en cumplimiento normativo sería conveniente que el consejo nombrase un miembro que se ocupara de manera continua de esta tarea o que se residenciara en alguna de sus comisiones. Igualmente, en la Política de cumplimiento IA sería oportuno que se estableciera que:

(a) el órgano de administración (o este miembro responsable) debe ser informado de cualquier incidencia grave del sistema de manera inmediata y tendrá una información contante y completa de los planes de repliegue;
(b) periódicamente el órgano de administración recibirá un informe de los riesgos que generan los sistemas de IA que produce la entidad;

(c) anualmente aprobará una política de "riesgos algorítmicos" en la que determinará los criterios de valoración del riesgo, señalará los controles generales que deben adoptarse, realizará las previsiones presupuestarías precisas para adoptarlos correctamente.

Cuando una empresa tiene como finalidad principal la producción de IA o es un responsable de despliegue (usuario profesional) que por el tipo de actividad que realiza puede estar sometido a reclamaciones de daños cuantiosas (vgr. un hospital) convendría que se dotara de una comisión de expertos independientes que analizará la idoneidad en abstracto de los controles; es decir, que estableciera un órgano de vigilancia con características similares a las que establece el art. 31 bis 2 2 ª del Código penal. Ello constituiría una "tercera línea de defensa" que aseguraría la efectividad de los controles y por tanto contribuiría al objetivo de contar con una IA más robusta. Sería importante que este órgano tuviera expresamente competencia para pronunciarse, a partir de su

independencia, acerca de si la inversión de medidas de seguridad y procedimientos de control es adecuada en atención a su costo y nivel de riesgos.

La pieza clave en el sistema de institucionalización es el encargado de IA – o si se prefiere *IA Compliance Officer* -. Se trata de un oficial de cumplimiento sectorial como el que quizás en entidades muy grandes puede existir en diversos ámbitos, como el lavado de activos o la protección de datos. Lógicamente requiere de unos conocimientos adecuados para realizar esta labor por lo que su perfil, mucho más técnico, será bien distinto. Dicho esto, lo que no son distintas sino muy similares son las funciones de que estará encargado que serán, fundamentalmente, las siguientes:

(a) supervisar que se ejecutan con regularidad los procedimientos que forman parte del sistema de gestión de la IA;

(b) asegurarse que los responsables de cada sistema realizan de manera continúa el análisis de riesgos;

(c) supervisar de manera intensa los procesos de retirada de sistemas y asegurarse la reparación de los daños la reparación;

(d) comunicar a las autoridades las incidencias graves y mantener la comunicación con ellas, asegurándose de que se cumplen sus instrucciones;

(e)comunicar al órgano de administración, o al administrador o comisión responsable, de manera inmediata las incidencias que hayan podido surgir y mantener una comunicación fluida con ellos;

(e) responsabilizarse de la alfabetización en IA y de la formación que deba impartirse a los responsables del despliegue;

(f) hacerse cargo de la interpretación y de las dudas que puedan surgir de la interpretación de los valores y normas de conducta que en materia de IA contiene el código ético;

(f) gestionar el sistema de documentación y trazabilidad;

(g) supervisar que los sistemas de IA que la organización adquiere sean adecuados y cumplan con las medidas de seguridad necesarias.

La primera línea de defensa es la que representan los dueños del control, esto es, los encargados directamente de ejecutar los procedimientos y controles. Cada sistema de IA generado debería tener, por ello, un responsable de seguridad, que informará de manera continua al IA Compliance officer y en casos especialmente graves al órgano de administración.

Como puede apreciarse, la institucionalización, la atribución de funciones y roles, requiere establecer un flujo de informaciones dentro de la entidad, que en este punto no tiene una estructura similar al que se prevé en otros muchos ámbitos y es desarrollado por las normas de estandarización. Mientras que el Reglamento IA no ha prestado

especial atención a los aspectos anteriores, si que se ha ocupado profusamente de los deberes de documentación en sus artículos 11 y 12.

II.- COMPLIANCE PROGRAMAS PARA SISTEMAS DE BAJO RIESGO

En este apartado final, a partir de cuanto se ha dicho estableceremos un programa de cumplimiento adecuado para el sistema APIA producto del presente proyecto de investigación. APIA no es un modelo de alto riesgo, lo que implica que de acuerdo con el Reglamento IA está sometido no tiene obligaciones legales de autorregulación; no hay autorregulación regulada sino voluntaria y tampoco es de uso general (art. 53 ss). El Reglamento anima a estos sistemas a adoptar los códigos de conducta que en el futuro publicará la Oficina de IA (art. 95), si bien pueden servir también de orientación la regulación de los sistemas de alto riesgo, siempre y cuando se tenga presente el principio de proporcionalidad del riesgo.

Para la organización de estos sistemas, pero en realidad también los productores de alto riesgo, debiera tenerse presente que en una organización es muy probable que coincidan los roles de productor/proveedor y responsable de despliegue, como usuario profesional. Aquí pueden existir combinaciones varias: por ejemplo, una universidad podría producir sistemas de bajo riesgo, pero ser usuario de despliegue de sistema de alto riesgo o, en ambos casos. El sistema de cumplimiento de IA de una organización debe ser lógicamente unitario y partir de todas estas situaciones.

Un aspecto importante es el que afecta a la contratación pública a la decisión acerca de los IA que deben adquirirse. Tal como se ha propuesto, el IA Compliance officer debiera aquí actuar juntamente con el órgano de contratación de la entidad, ya sea privada, pero sobre todo cuando sea pública.

Una organización, más allá de sus obligaciones legales en función del riesgo, también debiera tomar conciencia del grado de aceptación que existe entre sus usuarios de la IA, de la cultura de la organización en este punto. Una cultura adversa a la IA puede dar al traste con la innovación, por eso es necesario que emprenda campañas de alfabetización que no solo forme en IA, sino que además sirvan para animar a usuarios “acomplejados ante la tecnología” y para hacer una capacitación básica a todos los usuarios.

Otro elemento imprescindible es el código de conducta. De nuevo tomando como ejemplo una universidad habría que ocuparse de cuestiones tan simples como si un profesor puede corregir exámenes utilizando la IA o hasta qué punto resulta ético que los estudiantes la utilicen en sus actividades.

Para acometer todas estas tareas es indispensable una cierta institucionalización. El órgano de gobierno de la universidad debiera aprobar un código de conducta y nombrar un responsable de IA, que se ocupara de la formación, de tener noticia de todos los sistemas que pueden estar desarrollándose por los centros de investigación, de que se establecieran unos procedimientos de seguridad y desde luego que existieran los análisis de riesgos oportunos tanto en relación a los sistemas de que se producen, como los que se adquieren para ser utilizados, de manera profesional, o como consumidores. Entre los

valores a promocionar, y por tanto objeto de formación y protocolos, debiera estar el de la sostenibilidad y la inclusividad o no discriminación de usuarios.