ieee.es
Instituto Español de Estudios Estratégicos

Documento
Opinión

46/2019 28 de mayo de 2019

*Carlos Galán**

La certificación como mecanismo de control de la inteligencia artificial en Europa

La certificación como mecanismo de control de la inteligencia artificial en Europa

Resumen

La utilización de la inteligencia artificial (IA) constituye una de las más significativas aportaciones tecnológicas que impregnará la vida de las sociedades occidentales de los próximos años, en muchas de sus actividades cotidianas y en sus sectores más representativos, desde la industria al sistema financiero, pasando por la educación, la salud, el transporte y, desde luego, la defensa y la seguridad, aportando significativos beneficios, pero evidenciando también riesgos que es necesario valorar y minimizar.

Una realidad tan disruptiva como la IA exige que su tecnología y la de los productos y servicios sustentados en ella ofrezcan suficientes garantías de su adecuado funcionamiento.

El presente trabajo analiza y hace una propuesta para la aplicación de los mecanismos de Certificación de la Conformidad al mantenimiento de las antedichas garantías.

Palabras clave

Inteligencia artificial, certificación, control, legalidad y ética, defensa y seguridad.

*__NOTA:__ Las ideas contenidas en los ***Documentos de Opinión*** son responsabilidad de sus autores, sin que reflejen, necesariamente, el pensamiento del IEEE o del Ministerio de Defensa.

Certification as a control mechanism of Artificial Intelligence in Europe

Abstract

The use of Artificial Intelligence (AI) is one of the most significant technological contributions that will permeate the life of Western societies in the coming years, in many of its daily activities and in its most representative sectors, from industry to the financial system, going through education, health, transport and, of course, Defense and Security, providing significant benefits, but also showing risks that need to be assessed and minimized.

A reality as disruptive as AI requires that its technology and that of the products and services supported by it provide sufficient guarantees of its proper functioning.

The present work analyzes and makes a proposal for the application of the mechanisms of Certification of Conformity to the maintenance of the aforementioned guarantees.

Keywords

Artificial Intelligence, certification, control, legality and ethics, defense and security.

Introducción: la confiabilidad de la inteligencia artificial

El 27 de julio de 1987, el autor de este documento defendía ante el Tribunal de Tesis Doctoral de la Universidad Politécnica de Madrid el trabajo *Un modelo para la comprensión de problemas propuestos en lenguaje natural para un dominio limitado*. Como puede colegirse, la obra era el resultado de una investigación relacionada con uno de los paradigmas clásicos de la inteligencia artificial (IA): el reconocimiento automático del lenguaje natural.

Desde 1987, el reconocimiento del lenguaje natural, como el resto de los dominios competenciales de la IA, ha evolucionado enormemente.

Tras unos años de «travesía del desierto», prácticamente vacíos de los espectaculares resultados que se habían vaticinado, la IA goza en la actualidad de un magnífico presente, que derivará, con toda seguridad, en un extraordinario futuro[1].

No creemos pecar de quiméricos si afirmamos, con total rotundidad, que la IA constituirá uno de los ejes vertebradores de la sociedad de los próximos años; capaz de modificar nuestros hábitos y alterar nuestra relación con el mundo circundante[2]. No se trata solo de máquinas que juegan —y ganan— contra humanos, vehículos autónomos o robots que ayudan en tales o cuales tareas, etc. Estamos hablando de un nuevo ecosistema en el que se integrarán elementos inteligentes, capaces de razonar como lo haría un ser humano —en ocasiones, mejor y más rápidamente—. La velocidad de procesamiento de datos y de volumen manejado, obstáculos tradicionales de los desarrollos de la IA, se han derrumbado estrepitosamente. En la actualidad, ya es posible manejar en tiempo real terabytes de información (lo que ha hecho posible la aparición de nuevas herramientas como el *machine learning* o el *deep learning*) y, en general, la adopción de los métodos, procedimientos y herramientas de la IA en múltiples y variados escenarios: desde el vehículo autónomo, al cuidado de la salud, pasando por los robots domésticos o de servicio, la educación, la ciberseguridad[3], la medicina personalizada, la lucha contra

[1] Así lo sostiene también la UE en sus documentos: «Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions». *Coordinated Plan on Artificial Intelligence*. Brussels: 7/12/2018. COM 2018, pp. 795-final y «Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones». *Inteligencia artificial para Europa*. SWD 2018, pp. 137-final.

[2] Ello sin considerar las implicaciones económicas de la IA.

[3] La IA se encuentra cada vez más presente en la panoplia de herramientas usadas como mecanismo de defensa y ataque en el ciberespacio. Véase: *Informe de ciberamenazas y tendencias-Edición 2019*. Centro Criptológico Nacional; Centro Nacional de Inteligencia y *La inteligencia artificial aplicada a la defensa*. Documento de Trabajo 06/2018. Instituto Español de Estudios Estratégicos.

el cambio climático, el mejor uso de los recursos naturales o las infraestructuras de transporte, sin olvidar que, como se ha dicho, la IA se incluye claramente en ese grupo de tecnologías de marcado carácter dual, igualmente útil en aplicaciones para la defensa y la seguridad[4,5]. Todo ello en un escenario económico que podría superar los 38.000 millones de dólares en 2025[6].

Creemos que no está demasiado lejos el día en que las máquinas inteligentes, más allá de la realización de actividades rutinarias (lo que formaría parte de la denominada «IA débil»), serán capaces de mostrar comportamientos antes reservados al ser humano: la capacidad de razonar de forma autónoma, de aprender del pasado y de tomar decisiones (lo que se ha denominado «IA fuerte»).

Cuando alguno de mis alumnos me pregunta: «¿pueden pensar las máquinas?». Tras esbozar una sonrisa evocadora de Turing[7] o Minsky[8], respondo tajantemente que «una máquina, como usted las denomina, podría hacer lo mismo que cualquier ser humano, quien, por cierto, no es sino también una máquina. ¿O no?»[9].

Ante este futuro, más próximo de lo que pueda parecer, conviene estar preparados.

---

[4] ROLDÁN TUDELA, José Manuel. *La inteligencia artificial aplicada a la defensa*. Documento de Trabajo 06/2018. Instituto Español de Estudios Estratégicos.
La preocupación por la aplicación militar de la IA ha sido también señalada por MOLINER GONZÁLEZ, Juan A., que ha señalado: «Pero mientras los algoritmos de la inteligencia artificial avanzan imparablemente, los desafíos y retos éticos que se presentan en el combate deben ser analizados para que las máquinas no escapen al imprescindible control humano en su desarrollo y empleo, bajo adecuados principios morales».«Desafíos éticos en el uso militar de la inteligencia artificial», en *La Inteligencia Artificial aplicada a la Defensa*. IEEE, 2018, *op. cit.*

[5] Un buen ejemplo de la importancia conferida por los Estados nacionales a la IA de propósito militar lo constituyen los trabajos de la Agencia de Investigación del Ministerio de Defensa de los EE. UU. (DARPA), desarrollando en la actualidad importantes trabajos en tal sentido; o la actividad de la Agencia Europea de Defensa (EDA), con sus trabajos sobre robots para propósitos militares (proyecto MuRoC). La EDA presentaba la importancia de la IA en el anuncio de su proyecto «Artificial Intelligence and Big Data for Decision Making in C4ISR», de la siguiente forma: «Los Sistemas de Mando, Control, Comunicaciones, Computación e Inteligencia, junto con la Vigilancia y Reconocimiento (C4ISR), requieren de diversas tecnologías para, entre otras cosas, proporcionar una conciencia situacional y así poder apoyar en la toma de decisiones. Se requiere una aplicación innovadora de estas tecnologías para lograr la superioridad de la información que es crucial en las operaciones contemporáneas. En este contexto, hay evidencia emergente de que las tecnologías disruptivas de inteligencia artificial (IA) y big data (BD), en combinación con tecnologías de sensores e infraestructura más maduras, pueden ayudar a la comunidad de defensa a enfrentarse a los desafíos de los sistemas C4ISR contemporáneos, en términos de rendimiento, resiliencia, escalabilidad, interoperabilidad y eficiencia del operador». Tomado de DE LA FUENTE CHACÓN, José Carlos. «La inteligencia artificial y su aplicación en el mundo militar», en *La inteligencia artificial aplicada a la defensa.* IEEE, 2018, *op. cit.*

[6] Comité Económico y Social Europeo. 526.º Pleno del CESE, 31 de mayo a 1 de junio de 2017. Dictamen del Comité Económico y Social Europeo sobre la «inteligencia artificial: las consecuencias de la inteligencia artificial para el mercado único (digital), la producción, el consumo, el empleo y la sociedad».

[7] Disponible en https://www.csee.umbc.edu/courses/471/papers/turing.pdf.

[8] Disponible en https://www.muyinteresante.es/tecnologia/articulo/marvin-misnky.

[9] Ahorro al lector de la reproducción de los encendidos debates que suelen desencadenarse tras esta afirmación, en los que terminan apareciendo, inevitablemente, conceptos de orden filosófico, espiritual o religioso.

Y es aquí donde entra en juego un elemento de suma importancia, en cuya formulación encuentra su génesis el presente trabajo. Admitiendo que una máquina inteligente puede modificar su conducta atendiendo a su lógica interna y a factores exógenos, ¿quién nos asegura que su comportamiento va a desenvolverse dentro de lo que habitualmente denominamos «principios éticos»[10] y, en todo caso, respetando el ordenamiento jurídico vigente? ¿Qué tiene que decir la ley en esta situación? ¿Cómo ha de decirlo?

No se nos oculta que estas inquietudes alcanzan una cota máxima cuando se trata de aplicaciones de propósito militar. Como se ha dicho: «la comercialización de sistemas de IA con funciones más complejas está delegando la toma de decisiones en las mismas máquinas reduciendo la capacidad del ser humano en la toma final de la misma. No es extraño, por tanto, que se planteen dudas sobre los límites que deben establecerse para que esta intervención de los «algoritmos de IA» esté controlado»[11].

En el mismo sentido, el Comité Internacional de la Cruz Roja ha señalado: «la cuestión ética fundamental es si los principios de humanidad y los dictados de la conciencia pública pueden permitir que la decisión humana en el uso de la fuerza sea efectivamente sustituida con procesos controlados por computador, y las decisiones sobre vida y muerte sean cedidas a las maquinas»[12].

Ante este reto existen dos posibilidades no solo no excluyentes, sino complementarias: incorporando a los sistemas de construcción de tecnologías IA un conjunto de reglas capaces de ajustar el comportamiento de los productos o servicios IA a las normas ético-legales de aplicación; o asegurando, vía auditoría, que tales tecnologías, productos o servicios son conformes a un modelo o referencia de comportamiento ético-legal previamente definido.

Ambas respuestas son, como decimos, complementarias. El presente trabajo trata de la segunda de ellas: la certificación de conformidad (en base a auditorías independientes) como mecanismo de control ético-legal de las tecnologías, productos o servicios de la IA.

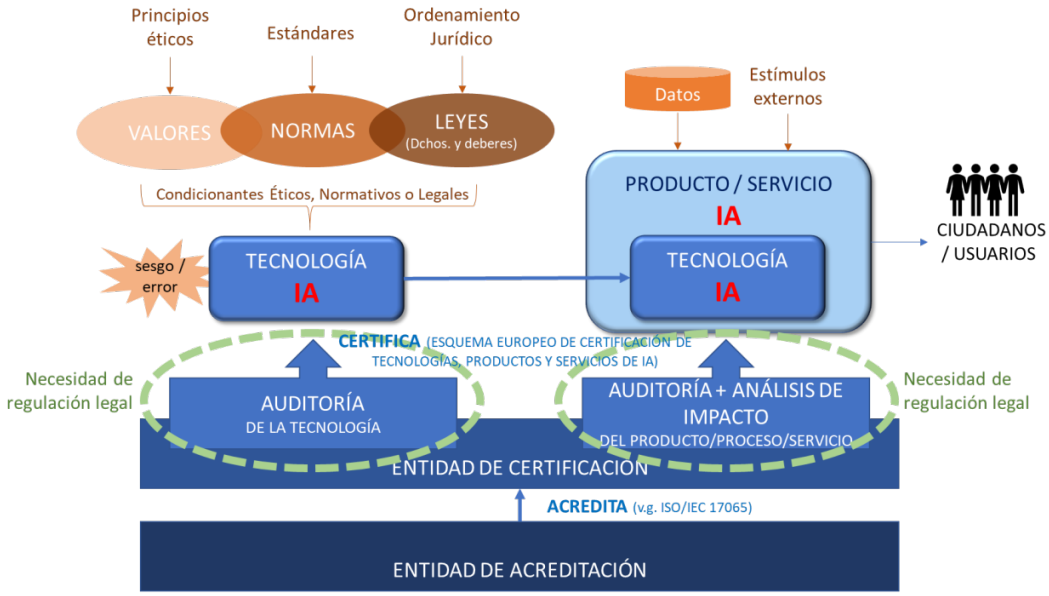
[10] Son varias las instituciones que han tratado de enumerar cuales deben ser tales principios éticos. Señalamos los propuestos por el Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías, de la Comisión Europea, en su *Declaración sobre inteligencia artificial, robótica y sistemas «autónomos»:* dignidad humana; autonomía; responsabilidad, justicia, equidad y solidaridad; democracia; estado de derecho y rendición de cuentas; seguridad, protección, e integridad física y mental; protección de datos y privacidad y sostenibilidad.

[11] ROLDAN TUDELA, José Manuel. *Op. cit.*

[12] ICRC (2018). *Ethics and autonomous weapon systems: An ethical basis for human control?* Ginebra.

El modelo propuesto

La figura siguiente muestra un esquema conceptual del modelo estudiado.



El modelo propuesto de Certificación de la IA en Europa

La tecnología, los productos y los servicios IA

Denominaremos «tecnología IA», en general y sin ánimo de exhaustividad, al conjunto de métodos, procedimientos, herramientas y resultados, científicos o tecnológicos, que constituyen la base para construir «productos o servicios IA», esto es, productos o servicios que poseen «capacidad para interpretar correctamente datos externos, aprender de tales datos y usar ese aprendizaje para lograr objetivos y tareas específicos a través de una adaptación flexible»[13].

La tecnología IA se configura, por consiguiente, como las piezas con las que construiremos soluciones sustentadas, en mayor o menor medida, en técnicas IA y cuyo comportamiento inteligente vendrá dado por la mayor o menor importancia o magnitud de la tecnología IA involucrada en el producto o servicio final.

Los productos/servicios IA, por su parte, son aquellos que podrán utilizar sus destinatarios, «consumidores» o usuarios finales, ya sean de «propósito general»

[13] No existe consenso sobre una definición formal de IA. Nos gusta y hemos utilizado la de Andreas Kaplan & Michael Haenlein: *Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*.

(dirigidos a todos) o de «propósito específico» (dirigidos a un sector de la población o de la industria), y entre cuyos componentes configuradores se encuentran elementos pertenecientes a la tecnología IA. Por ejemplo, un vehículo autónomo (paradigma de producto IA), dispondrá de un motor de inferencia (tecnología IA) encargado de procesar los estímulos del exterior —y del interior, en su caso— conforme a lo dispuesto en el *software* IA utilizado, atendiendo a las señales que recibe de los sensores y, seguramente, a la experiencia acumulada en base al tratamiento previo de miles o millones de datos.

Por tanto, no todos los elementos constituyentes de un producto IA han de ser, obligatoriamente, tecnología IA. Por el contrario, lo habitual es que en un producto/servicio IA coexistan elementos desarrollados e implementados usando lo que podemos denominar «tecnologías tradicionales» con componentes desarrollados en base a tecnología IA.

Los elementos configuradores de la tecnología IA y los productos/servicios IA

Mientras que la «tecnología tradicional» se sustenta en el diseño e implementación de «algoritmos deterministas», esto es, procedimientos en los que para unos mismos datos de entrada se obtienen los mismos resultados. Los modelos de IA (muy especialmente los comprendidos en la denominada IA fuerte) utilizan «procedimientos heurísticos», es decir, modelos computacionales que no solo tienen en cuenta los datos de entrada, sino también la experiencia acumulada y el conocimiento derivado de tal experiencia, lo que facilita que, frente a los mismos estímulos, puedan generarse respuestas diferentes. Esto es lo que denominaremos «capacidad adaptativa de la IA», cualidad que acerca estos sistemas al razonamiento humano.

Este comportamiento heurístico es el que confiere a la IA sus mayores beneficios y, también, sus más significativos riesgos.

Que una máquina inteligente —un sistema IA, en definitiva— sea capaz de derivar su funcionamiento autónomo hacia comportamientos poco éticos o ilegales solo depende de que sus desarrolladores hayan incorporado a la tecnología IA usada un cierto tipo de elementos, fiables y contrastados, que impidan tal deriva[14].

[14] En el terreno de las aplicaciones militares conviene mencionar –como acertadamente ha señalado MOLINER GONZÁLEZ, Juan A. IEEE, 2018, *op. cit*.– los trabajos de ARKIN, Ronald. *Ethical Robots in Warfare*. IEEE Technology and Society Magazine. Spring 2009, cuyo principal postulado es que «los sistemas de IA puedan programarse con determinadas restricciones que salvaguardarían el respeto a las

He aquí la esencia del problema: ¿cuáles son esos elementos capaces de hacer que un sistema IA funcione adecuadamente conforme a los principios éticos y legales asumidos por la comunidad de usuarios al que va dirigido?, y ¿cómo incorporar tales principios en el sistema?[15]

Son varios los aspectos que pueden —y deben— dirigir el diseño y desarrollo de tecnología IA, y que deberían ser tenidos en cuenta por sus fabricantes.

Los más significativos son los mostrados en el cuadro siguiente:

Aspectos de carácter exógeno a la tecnología IA	Las normas jurídicas	Conjunto de regulaciones pertenecientes al ordenamiento jurídico aplicable.
	Los principios éticos	Conjunto de directrices de comportamiento, generalmente aceptadas por la comunidad destinataria de los productos/servicios IA[16].

reglas éticas y al derecho internacional humanitario en el campo de batalla sin el riesgo del fallo humano que puede llevar al acto ilegal y, sobre todo, inmoral, en el desarrollo de las operaciones militares».

[15] Especialmente, cuando es la propia Comisión Europea la que pretende facilitar el acceso a las últimas tecnologías a todos los usuarios potenciales: las pequeñas y medianas empresas, las empresas de sectores no tecnológicos y las Administraciones Públicas, alentándoles a ensayarlos. «Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones». *Inteligencia artificial para Europa*.

[16] Los «principios éticos» pueden formularse desde distintos puntos de vista. El más habitual, no obstante, es hacerlos coincidir con el respeto a los derechos formulados en las cartas internacionales de derechos humanos, desde la Declaración Universal de Derechos Humanos de 1948, hasta nuestra Constitución de 1978, pasando por los tratados de la UE, la Carta de los Derechos Fundamentales de la UE, el Convenio Europeo de Derechos Humanos o, incluso, la Carta Social Europea o el Reglamento General de Protección de Datos.

El *Artificial Intelligence High-Level Expert Group* europeo (AI HLEG) ha señalado la relación entre «principios-valores-derechos fundamentales» del siguiente modo. «Los derechos fundamentales proporcionan la base para la formulación de los principios éticos. Esos principios son normas abstractas de alto nivel que los desarrolladores, implementadores, usuarios y reguladores deben seguir para defender el propósito de una IA centrada en el hombre y de confianza. Los valores, a su vez, proporcionan una orientación más concreta sobre cómo defender los principios éticos, al tiempo que respaldan los derechos fundamentales». AI HLEG, *Ethics Guidelines for Trustworthy AI*. Draft-Dec., 2018. Para el propósito del presente trabajo, hablaremos, genéricamente, de «principios éticos».

Otras referencias bibliográficas en torno a los principios que deben conducir la IA pueden encontrarse en los siguientes textos: *Asilomar AI Principles*, desarrollados por el Future of Life Institute (2017); *Montreal Declaration for Responsible AI*, de la Universidad de Montreal (2017); la segunda versión de los principios generales del IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems* (2017), los *Ethical Principles* del Grupo de Ética de la Ciencia y las Nuevas tecnologías de la Comisión Europea (2018); los *Five overarching principles for an AI code,* recogidos en el §417 del Informe del Comité para la Inteligencia Artificial de la Cámara de los Lores británica (2018); y los *Tenets of the Partnership on AI* (2018).

	Los estándares	Conjunto de normas técnicas, generalmente provenientes de organizaciones internacionales[17], dirigidas a normalizar la construcción o el uso de productos, procesos o servicios[18].
Aspectos de carácter endógeno a la tecnología IA	El sesgo	Cualidad indeseable —habitualmente provocada por un incorrecto diseño o una insuficiente o inadecuada elección de los datos de aprendizaje— que puede hacer que los sistemas IA produzcan resultados sesgados.
	El error	Fallos —habitualmente, no deliberados— en el diseño o el desarrollo de la tecnología.

Todos ellos conforman los elementos éticos, normativos o legales que condicionan la construcción de la tecnología IA. Si alguno no está presente o está inadecuadamente implementado, la tecnología IA resultante no poseerá las debidas garantías que aseguren que su funcionamiento se ajusta a los modelos legales o éticos deseados por la comunidad usuaria de los productos o servicios finales.

Salvo que posea barreras *ad hoc* —cosa que no suele ser frecuente—, el comportamiento de la tecnología IA, bueno o malo, aceptable o no aceptable, se traslada, querámoslo o no, al producto/servicio IA en el que se inserta. Como quiera que es precisamente este producto o servicio el utilizado por los usuarios finales, es en ese momento cuando se evidencia la idoneidad de la tecnología de base utilizada, y cuando, desafortunadamente, la remediación suele ser difícil –o imposible, incluso– si tal comportamiento se ha alejado de lo lícito o lo deseable.

[17] Las normas más significativas provienen de ISO (*International Standarization Organization*), IEC (*International Electrotechnical Commission*), IEEE (*Institute of Electrical and Electronics Engineers*), ITU (International Telecommunications Union) o, de alcance europeo, ETSI (*European Telecommunications Standards Institute*).

[18] Aunque, en puridad, los estándares forman parte de los que hemos denominado «aspectos exógenos», su generalizada utilización práctica los ha venido convirtiendo, de facto, en elementos consustanciales al desarrollo de los algoritmos, más cerca, en muchas ocasiones, a los componentes «endógenos» de los sistemas.

Control ex-ante y control ex-post

Así las cosas, parece lógico pensar que, antes que un producto/servicio IA se introduzca en el mercado, es necesario asegurar que su comportamiento final será el esperado. Este aseguramiento pasa, necesariamente, por garantizar la idoneidad del propio producto/servicio, así como la tecnología usada para su construcción. En otras palabras, aunque constituye la verificación más importante, no basta con garantizar que una tecnología IA es adecuada desde los puntos de vista ético y legal, es necesario, además, verificar que el producto/servicio IA en el que se «inserta» es igualmente aceptable tras dicha inclusión[19].

<table>
<tr><td>Control ex-ante:
medidas de control <u>previas</u> al desarrollo o implementación de la tecnología, producto o servicio IA.</td><td>Seguridad desde el diseño.
Actividades esenciales comprendidas:
- Concepción.
- Planificación.
- Desarrollo.
- Pruebas.</td><td rowspan="2">Verificación de la conformidad de la solución con:
• El ordenamiento jurídico aplicable.
• Principios éticos.
• Estándares técnicos.
• Análisis de impacto en la sociedad (comunidad destinataria de los productos/servicios)[20].

Satisfacción de las exigencias de:
• Transparencia e inteligibilidad de los sistemas.
• Posibilidad de acceso y verificación.</td></tr>
<tr><td>Control ex-post:
Medidas de control <u>posteriores</u> al desarrollo o implementación de la tecnología, producto o servicio IA.</td><td>Seguridad en la operación.
Actividades esenciales comprendidas:
- Integración.
- Adquisición.
- Despliegue.</td></tr>
</table>

[19] Una tecnología IA, originariamente «inocua», insertada en un producto o servicio inadecuadamente configurado o protegido podría dar como resultado un producto/servicio no deseable.

[20] En base a los trabajos de Stahl, Timmermans y Flick (*Ethics of Emerging Information and Communication Technologies*), podemos distinguir los problemas que podrían tener impacto a nivel individual (como la autonomía, la identidad, la dignidad, la privacidad y la protección de datos) y los que tienen un impacto a nivel social (como la imparcialidad y la equidad, la identidad colectiva y el estado del bienestar, la responsabilidad, la rendición de cuentas y la transparencia, la privacidad en relación con la vigilancia, la democracia y la confianza).

	- Explotación. - Publicación. - Conservación. - Acceso. - Interconexión.	• Explicabilidad, trazabilidad y rendición de cuentas (responsabilidad).

La verificación de dicha idoneidad puede hacerse en dos momentos temporales: antes de su construcción o tras ella. Ambos tipos de control no son excluyentes.

El cuadro anterior muestra las características esenciales de ambos tipos de control[21].

Aun siendo siempre preferible, el control ex-ante no es fácil de implantar. Como hemos señalado en el cuadro anterior, exige de los fabricantes de la tecnología, productos o servicios IA (y, en su caso, de sus distribuidores) el sometimiento a una disciplina de actuación que ha de impregnar no solo la concepción de los sistemas sino su completo ciclo de vida: planificación, diseño, adquisición, construcción, despliegue, explotación, publicación, conservación y acceso o interconexión con los mismos.

Por otro lado, un control ex-ante exige también disponer de mecanismos fiables, capaces de trasladar al diseño y desarrollo de la tecnología IA subyacente a las exigencias legales, éticas y normativas que aseguren que, a la postre, se obtiene un producto adecuado, jurídica y éticamente aceptable[22].

En este sentido han ido dirigidos los trabajos del Grupo de Expertos de Alto Nivel en Inteligencia Artificial de la Comisión Europea (AIHLE) que ha insistido en señalar, acertadamente, que la IA debe estar centrada en el ser humano: «la IA debe desarrollarse, desplegarse y utilizarse con un "propósito ético", fundamentada en los derechos fundamentales, los valores sociales y los principios éticos de beneficencia

[21] El Real Decreto 4/2010, de 8 de enero, por el que se regula el Esquema Nacional de Interoperabilidad ya recogía la preocupación por extender la garantía de seguridad a todo el ciclo de vida de los sistemas de información, al expresar, en su artículo 1.2: «2. El Esquema Nacional de Interoperabilidad comprenderá los criterios y recomendaciones de seguridad, normalización y conservación de la información, de los formatos y de las aplicaciones que deberán ser tenidos en cuenta por las Administraciones Públicas para asegurar un adecuado nivel de interoperabilidad organizativa, semántica y técnica de los datos, informaciones y servicios que gestionen en el ejercicio de sus competencias y para evitar la discriminación a los ciudadanos por razón de su elección tecnológica».

[22] No es necesario para el propósito de este trabajo distinguir entre los conceptos *hard-ethics* y *soft-ethics*, tal y como han sido definidos, por ejemplo, por Luciano Floridi en *Soft Ethics and the Governance of the Digital*.

(hacer el bien), no-maleficencia (no hacer daño), la autonomía del ser humano, la justicia y la explicabilidad[23]. Todo ello es crucial para lograr una IA confiable»[24].

Desde el Comité Económico y Social Europeo se ha señalado que la IA debe responder a unos estándares mínimos de seguridad (interna y externa) que le permitan funcionar correctamente y sin producir daños a los usuarios o destinatarios de su actuación, llegando a proponer, incluso, la elaboración de un código deontológico, uniforme y universal, para el desarrollo, despliegue y utilización de la IA, de modo que durante todo su proceso de funcionamiento los sistemas de IA sean compatibles con los principios de la dignidad humana, la integridad, la libertad, la privacidad, la diversidad cultural y de género y los derechos humanos fundamentales[25,26].

Además, en nuestra opinión, para que una tecnología, producto o servicio IA pueda ser usada por sus destinatarios, no es únicamente imprescindible que sea conforme a unos principios éticos o a un ordenamiento jurídico concreto; será necesario, adicionalmente, realizar un análisis de impacto de tal objeto IA sobre la sociedad en la que pretende insertarse. Esto es especialmente importante cuando el despliegue del objeto IA de que se trate pueda tener un impacto significativo en los derechos y libertades de las personas o en el orden socio-económico establecido[27].

A la vista de la complejidad para trasladar los principios éticos y legales a la tecnología IA (y, en su consecuencia, a los productos o servicios en los que se incorpora), parece necesario contemplar, complementariamente, un control ex-post, capaz de determinar, con un alto grado de confianza, la bondad de un producto/servicio IA, ya construido, y antes de su comercialización o despliegue.

---

[23] Término que debemos interpretar como la capacidad para explicar las conclusiones alcanzadas y su trazabilidad.

[24] A fecha de la redacción de este trabajo, el High-Level Expert Group on Artificial Intelligence ha publicado un borrador de *Ethics Guidelines for Trustworthy AI* (diciembre, 2018), antesala del futuro documento final.

[25] Dictamen CESE, *op. cit.*

[26] El Parlamento Europeo, en sus *Normas de derecho civil sobre robótica* [Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de derecho civil sobre robótica (2015/2103 INL)], ha incluido un *Código de conducta ética para los ingenieros en robótica*, cuyos postulados podrían ser en parte trasladables a un eventual código ético de la IA. Por otro lado, la necesidad de contar con un Código de conducta ética, como instrumento de guía para los desarrolladores de IA también ha sido puesto de manifiesto en *Towards a Global Artificial Intelligence Charter*. (METZINGER, Thomas. Contenido en *Should we fear artificial intelligence?* European Parliament, 2018).

[27] VIDA, José; señala: «… las iniciativas que se han desarrollado hasta ahora recomiendan aumentar y profundizar el conocimiento sobre la IA en todos los niveles para poder reconocer, definir y controlar las disrupciones en su desarrollo a fin de poder regularlas adecuadamente y a su debido tiempo». «Los retos de la regulación de la Inteligencia Artificial: Algunas portaciones desde la perspectiva europea», en *Sociedad Digital y Derecho*. BOE, Nov. 2018.

Auditoría y Certificación de Conformidad

Como es sabido, una auditoría es «un proceso sistemático, independiente y documentado que persigue la obtención de evidencias objetivas y su evaluación para determinar en qué medida se cumplen los criterios de auditoría»[28]. Dicho en otras palabras, una auditoría señala hasta qué punto un determinado objeto es conforme con lo dispuesto en la norma de referencia de que se trate.

Este tipo de controles ex-post permiten concentrar nuestra atención en la observación del comportamiento del objeto auditado, más que en asegurar que tal objeto ha sido diseñado o construido conforme a unas determinadas reglas. Se trata, por tanto, de una evaluación basada en las evidencias que se desprenden del análisis objeto auditado, cuando se le somete a una adecuada batería de controles, y que nos permitirán decidir si tal objeto se encuentra dentro de los márgenes admitidos por la norma de referencia. En otras palabras, y acercándolo a nuestro propósito, si su comportamiento está alineado con las expectativas éticas y legales que dirigen la comunidad en la que pretende integrarse.

Si esta «Auditoría de Conformidad» resulta satisfactoria, la Entidad de Evaluación de la Conformidad (por sí misma o por medio de un tercero habilitado) expedirá una «Certificación de Conformidad» que exhibirá, *erga omnes*, la conformidad de la tecnología, producto o servicio con la norma (esquema de certificación) que se ha tomado como referencia[29].

La utilización de las certificaciones, basadas en auditorías independientes, como elemento de exhibición de la conformidad de un determinado producto o servicio es algo conocido. La conformidad con SOG-IS MRA[30], los modelos de certificación de los Servicios de Confianza[31], los esquemas de certificación de cumplimiento del RGPD, o

[28] *UNE-EN ISO 19011:2018 Guidelines for auditing management systems.*

[29] Esta parece ser también la pretensión del Comité Económico y Social Europeo, que, en la obra citada, señala textualmente: «El CESE aboga por una infraestructura de IA europea de fuente abierta (open source), que incluya entornos de aprendizaje respetuosos de la vida privada, entornos de ensayo en condiciones reales (*real life*) y conjuntos de datos de alta calidad para el desarrollo y la formación de sistemas de IA. El CESE destaca la ventaja (competitiva) que puede obtener la UE en el mercado mundial mediante el desarrollo y la promoción de «sistemas de IA de responsabilidad europea, provistos de un sistema europeo de certificación y etiquetado de la IA», añadiendo: «en este sentido, se propone la utilización de un sistema de normalización para la verificación, validación y control de los sistemas de IA, basado en un amplio espectro de normas en materia de seguridad, transparencia, inteligibilidad, rendición de cuentas y valores éticos».

[30] Senior Officials Group-Information Systems Security (SOG-IS) Mutual Recognition Agreement (MRA) of Information Technology Security Evaluation Certificates.

[31] Reglamento (UE) 910/2014 del Parlamento Europeo y del Consejo, de 23 de julio de 2014, relativo a la identificación electrónica y los servicios de confianza para las transacciones electrónicas en el mercado interior y por la que se deroga la Directiva 1999/93/CE.

los modelos de certificación de la seguridad auspiciados por la denominada *Cibersecurity Act*[32] son buenos ejemplos. En España, el modelo más extendido para evaluar la conformidad de la seguridad de los sistemas de información (especialmente dirigido a las entidades del Sector Público y a las empresas privadas proveedoras de servicios a aquellas) lo constituye el Esquema Nacional de Seguridad[33], que contempla la Certificación de Conformidad con el ENS en base a la superación de auditorías de cumplimiento periódicas por parte de los sistemas de información concernidos; o el modelo de Evaluación y Certificación de la Seguridad de las Tecnologías de la Información[34].

En nuestro caso, para hacer posible este mecanismo de certificación es necesario contar con dos elementos previos:

1. Un Esquema de Evaluación y Certificación de la Conformidad de las Tecnologías, productos o servicios IA[35].
2. Una norma europea (de tipo reglamentario, preferentemente) que regule el esquema anterior, su desarrollo, aplicación y actualización; y que determine los actores involucrados en su despliegue: Entidades de Acreditación y Entidades de Evaluación de la Conformidad, esencialmente[36].

Es evidente que la mayor dificultad de este modelo se encuentra en el primero de los puntos citados y es ahí donde, a nuestro juicio, deberían centrarse los primeros trabajos. En nuestra opinión, solo debería desplegarse en Europa aquella tecnología, producto o servicio IA que se encuentre debidamente certificado.

[32] Proposal for Regulation of the European Parliament and of the Council on ENISA, the «EU Cybersecurity Agency», and repealing Regulation (EU) 526/2013, and on Information and Communication Technology cybersecurity certification ("Cybersecurity Act")».

[33] Real Decreto 3/2010, de 8 de enero, por el que se regula el Esquema Nacional de Seguridad (ENS).

[34] Orden PRE/2740/2007, de 19 de septiembre, por la que se aprueba el Reglamento de Evaluación y Certificación de la Seguridad de las Tecnologías de la Información.

[35] Pese a los indudables beneficios derivados de los modelos de certificación, no debemos olvidar, no obstante, que una certificación no puede garantizar totalmente que un producto o servicio TIC sea completamente seguro. Así lo recuerda la Cibersecurity Act citada, en su defensa de los esquemas de certificación de tecnologías, productos y servicios de ciberseguridad.

[36] La necesidad regulatoria también está en la Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Inteligencia artificial para Europa, cuando señala: «si bien la autorregulación puede proporcionar un primer conjunto de índices de referencia con respecto a los cuales sea posible valorar las aplicaciones y resultados que van apareciendo, las autoridades públicas deben garantizar que los marcos reglamentarios para el desarrollo y el uso de las tecnologías de IA estén en consonancia con esos valores y derechos fundamentales».

Conclusiones y acciones subsiguientes

De todo lo anterior, podemos extraer algunas conclusiones y proponer acciones subsiguientes:

- Siendo la IA un elemento indispensable para el desarrollo de la sociedad, es necesario garantizar, hasta donde sea posible, que la tecnología IA, los productos IA y los servicios IA se ajustan a lo dispuesto en la regulación legal que resulte de aplicación, manteniendo un escrupuloso respeto a los principios éticos universalmente aceptados.
- La garantía anterior solo puede alcanzarse con una adecuada combinación de dos tipos de controles: ex-ante y ex-post, antes y después, respectivamente, del diseño, desarrollo, implementación o despliegue de la tecnología, productos o servicios implicados.
- El control ex-ante, aunque presenta indudables y lógicas ventajas, es complejo y está sometido, en cualquier caso, a la buena voluntad de los diseñadores, desarrolladores o implementadores.
- El control ex-post, materializado en forma de auditorías independientes y certificaciones de conformidad, se configura como un mecanismo práctico y eficaz para alcanzar los objetivos perseguidos.
- La implantación de las auditorías de conformidad IA exige disponer de un Esquema Europeo de Certificación de Tecnologías, Productos y Servicios IA[37]; así como un elenco de entidades de evaluación de la conformidad que dispongan de las capacidades técnicas requeridas, la independencia y la imparcialidad debidas, y hayan sido debidamente habilitadas por una entidad de acreditación[38].
- Todo ello deberá estar perfectamente regulado a través de la normativa europea correspondiente[39].
- Las tecnologías, productos o servicios IA pueden construirse dentro de la UE o provenir de terceros países. La regulación anterior debe aplicarse a cualquier

[37] Actividad que muy bien podría ser desarrollada a través del *European Cybersecurity Certification Group.*
[38] Que será el organismo nacional de acreditación designado de conformidad con el Reglamento (CE) 765/2008 del Parlamento Europeo y del Consejo, de 9 de julio de 2008, por el que se establecen los requisitos de acreditación y vigilancia del mercado relativos a la comercialización de los productos y por el que se deroga el Reglamento (CEE) 339/93 (DO L 218 de 13.8.2008, p. 30), por ejemplo, con arreglo a la norma EN ISO/IEC 17065/2012.
[39]«La traslación de los valores y las normas en el diseño y funcionamiento de los sistemas IA deben formar parte de los marcos regulatorios». *Artificial Intelligence. A European Perspective.* Joint Research Centre (JRC), the European Commission's science and knowledge service.

tecnología, producto o servicio de este tipo que pretenda desplegarse o comercializarse en Europa. Solo la tecnología, los productos o los servicios IA certificados deberían beneficiarse de ayudas públicas[40].

- Los esquemas de certificación de tecnologías, productos y servicios IA deben incorporarse a los planes derivados de la Estrategia Española de I+D+i en Inteligencia Artificial.

De la misma forma, en el entorno de defensa y seguridad, el empleo de la IA es igualmente un tema sensible por las implicaciones éticas y legales que trae consigo, siendo necesario, como así se ha dicho, «establecer unos conceptos de empleo que aseguren, en lo posible, que la aplicación militar de estas tecnologías sea aceptable social y legalmente, tanto a nivel nacional como internacional»[41].

Reiterando la importancia de la IA en el futuro de las sociedades occidentales y la perentoria necesidad de asegurar su adecuado despliegue, nos gustaría terminar con una frase extraída de la ya citada Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones sobre una inteligencia artificial para Europa: «nuestra forma de abordar la cuestión de la IA definirá el mundo en el que vamos a vivir».

*Carlos Galán**
Licenciado en Derecho y abogado especialista en Derecho de las TIC
Doctor en Informática

[40] Como señala la Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. *Inteligencia artificial para Europa: «*El principio rector de todas las ayudas para la investigación en materia de IA será el desarrollo de una "IA responsable", centrada en el ser humano; véase la línea de trabajo de la Comisión «Investigación e Innovación Responsables». Disponible en: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation

[41] ROLDÁN TUDELA, José Manuel. «La inteligencia artificial y la fricción de la guerra», en *La inteligencia artificial aplicada a la defensa,* IEEE, 2018, *op. cit.,* que añade*:* «La IA y la RI cambiarán el carácter de la guerra, como ha ocurrido en diversas ocasiones a lo largo de los siglos. Los cambios que introducirán estas tecnologías pueden ser profundos, ya que incorporan capacidades que superan no solo la aptitud física del ser humano, sino también parte de sus facultades mentales».