

Capítulo XXIII

Análisis comparado de las definiciones jurídicas de «sistema de inteligencia artificial», ¿es necesaria una armonización normativa internacional?

Santiago Robert Guillén¹

1. INTRODUCCIÓN: INTELIGENCIA, ¿ARTIFICIAL?

Hasta la fecha, no se ha logrado una definición ampliamente aceptada del concepto de «inteligencia», lo que pone de manifiesto las dificultades para definir «inteligencia artificial» (Pino Diez et al., 2001). El uso de este término puede generar expectativas poco realistas sobre las tecnologías actuales (Mirzadeh et al., 2024). Por ello, las normativas emergentes evitan definir explícitamente «inteligencia artificial» y, en su lugar, optan por caracterizar a los «sistemas de inteligencia artificial» a través de la descripción de sus funciones operativas y características, mostrando que, aunque algunos resultados generados por estos sistemas pueden ser comparables a los obtenidos por procesos cognitivos humanos, los procedimientos que siguen son fundamentalmente distintos, ya que se basan en algoritmos predefinidos y carecen de autoconciencia y experiencias subjetivas (Franceschetti, 2018; Floridi et al., 2019). Los «sistemas de inteligencia artificial» comprenden programas informáticos, aplicaciones y dispositivos que implementan tecnologías también denominadas «de inteligencia artificial», como redes neuronales (*neural networks*), algoritmos de aprendizaje profundo (*deep learning algorithms*) y procesamiento del lenguaje natural (*natural*

1. Profesor de la Universidad Autónoma de Barcelona (UAB) y Doctor en Derecho por esta misma institución universitaria (<https://orcid.org/0000-0002-6919-6801>). Este estudio se enmarca en el Proyecto de investigación I+D+i «EdU-InA: Políticas y Prácticas sobre la Inteligencia Artificial Generativa en la educación universitaria» (PID2023-149069OA-I00).

language processing), y que son implementaciones concretas de principios y teorías que sustentan el campo interdisciplinario de la «inteligencia artificial» (LeCun et al. 2015; Jurafsky et al., 2024), lo que añade una capa adicional de complejidad y múltiples interpretaciones al concepto. No obstante, desde una perspectiva pragmática y regulatoria, resulta más efectivo utilizar expresiones como «sistemas de inteligencia artificial» o «sistemas basados en inteligencia artificial» acompañadas de descripciones funcionales que establezcan criterios medibles para evaluar la conformidad de estos sistemas, que utilizar otras excesivamente técnicas y complejas que, con toda probabilidad, tampoco resultarían definitivas, y que podrían desencadenar debates terminológicos interminables (Pino Diez et al., 2001; Floridi et al., 2019). A pesar de este enfoque pragmático, el desafío de actualizar aquellas *descripciones operativas* de los sistemas es ineludible, a fin de asegurar que evolucionen en consonancia con los avances tecnológicos en un campo que, por su naturaleza, está en constante desarrollo y cuya complejidad aún no se comprende del todo (Grieman, 2023). Sin embargo, este reto también se manifiesta en la falta de consenso sobre qué funciones específicas deben definir un «sistema de inteligencia artificial», generando incertidumbres tanto en el ámbito académico como en el normativo. Además, la aplicación coherente de las normativas a nivel internacional puede verse obstaculizada no solo por las discrepancias en la comprensión tecnológica, sino también por divergencias lingüísticas y disparidades en las prioridades regulatorias entre distintas jurisdicciones (Cath et al., 2018). Dada la naturaleza dinámica y la evolución acelerada de esta tecnología, parece necesario alcanzar un consenso internacional que incluya una definición unificada y aplicable de «sistemas de inteligencia artificial», para abordar los desafíos transfronterizos que plantea, asegurar la coherencia regulatoria y efectividad de las normativas, sin obstaculizar la innovación (Floridi et al., 2018; Wallach et al., 2018). Como señala el Considerando 12 del Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, que establece normas armonizadas en materia de inteligencia artificial en la Unión Europea (Diario Oficial de la Unión Europea (DOUE) núm. 1689, de 12 de julio de 2024), «*debe definirse con claridad el concepto de “sistema de IA” en el presente Reglamento y armonizarlo estrechamente con los trabajos de las organizaciones internacionales que se ocupan de la IA, a fin de garantizar la seguridad jurídica y facilitar la convergencia a escala internacional y una amplia aceptación, al mismo tiempo que se prevé la flexibilidad necesaria para dar cabida a los rápidos avances tecnológicos en este ámbito*».

A partir de esta premisa, y con base en el mencionado Reglamento europeo, analizaré la evolución de la definición jurídica de «sistema de inteligencia artificial» (en

adelante, sistema de IA) dentro de este marco hasta su versión final, y la compararé con otras regulaciones fuera de la UE, para determinar si un sistema de IA debe entenderse únicamente como *software* o si puede incluir otros componentes (II.); el rol de la autonomía y las técnicas avanzadas en su definición jurídica (III.); las dificultades lingüísticas y conceptuales que surgen al traducir las definiciones en distintos idiomas (IV.); y un análisis estructural de los sistemas, centrado en los modelos, los procesos de inferencia y la terminología técnica (V.). Finalmente, concluiré con una propuesta de contenido de contenido mínimo para una eventual definición y una reflexión sobre la necesidad de alcanzar un consenso global en la regulación (VI.).

2. LA NATURALEZA DE LOS SISTEMAS DE IA: ¿SON SOLO SOFTWARE O ALGO MÁS?

La Comisión Europea, en su Propuesta de Reglamento sobre Inteligencia Artificial, de 21 abril de 2021 (COM(2021) 206 final), subrayó la necesidad de establecer «*con claridad la noción de sistema de IA para ofrecer seguridad jurídica*» (Considerando 5) y definió un sistema de IA como «*el software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa*». Esta definición difiere de manera sustancial de la que figura en el Documento de Trabajo del Personal de la Comisión (*Impact Assessment*, SWD(2021) 84 final) que acompaña la propuesta, el cual, siguiendo la definición adoptada en 2019 por la Organización para la Cooperación y el Desarrollo Económico (OCDE) en su *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449, 2019), describe un sistema de IA como «*a machine-based system that can, for a given set of human-defined objectives, generate output such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy*». El Documento de Trabajo, de carácter no vinculante, cumple una función de apoyo y análisis, proporcionando un marco conceptual más amplio. No obstante, la Comisión, sin apartarse de este enfoque general, adoptó en su propuesta una definición más específica, centrada únicamente en el software, probablemente orientada a garantizar una mayor precisión regulatoria, pero que excluye otros componentes técnicos que también podrían formar parte de un sistema de IA, como se analizará a continuación.

La definición de la Comisión se limita, por lo tanto, al componente *algorítmico* de los sistemas de IA, es decir, al conjunto de instrucciones codificadas en lenguaje de programación (código fuente) y a los modelos matemáticos subyacentes. Estos *modelos*, entendidos como el conjunto de reglas y algoritmos que utiliza el sistema para procesar datos, extraer patrones y generar predicciones o decisiones (Koltsakis et al., 2023), incluyen, entre otras técnicas, el aprendizaje profundo (*deep learning*), el procesamiento del lenguaje natural (NLP, por sus siglas en inglés) así como redes neuronales y los algoritmos de optimización (Franceschetti, 2018; Floridi, 2021). Desde una perspectiva jurídica, el software presenta la ventaja práctica de ser un objeto delimitable, verificable y susceptible auditorías técnicas. No obstante, los sistemas basados en aprendizaje profundo plantean retos adicionales en términos de transparencia y auditabilidad, debido a la naturaleza opaca de sus procesos de decisión (Cheong, 2024). En cualquier caso, aunque un sistema de IA puede estar fundamentalmente basado en software, también existen sistemas de naturaleza *integrada* que incorporen componentes de hardware indispensables para su funcionamiento (Radanliev et al., 2021; Talib et al., 2021). Por ejemplo, en dispositivos médicos como el *Da Vinci Surgical System*, el software es esencial para ejecutar los algoritmos, pero el hardware (brazos robóticos, sensores, etc.) es igualmente indispensable para su operatividad. Por tanto, la definición propuesta por la Comisión podría ser incompleta al omitir aspectos clave del funcionamiento de estos sistemas. Este punto no pasó inadvertido para el Consejo de la Unión Europea (en adelante, Consejo de la UE), que orientó su propuesta de definición en un esfuerzo por capturar las distintas configuraciones tecnológicas de IA, tanto en el ámbito digital como físico, como veremos. De hecho, en 2018, la propia Comisión Europea, mediante de su *Comunicación sobre la inteligencia artificial para Europa* (COM(2018) 237 final), presentó una descripción de sistema de IA que incluía tanto software como hardware, destacando que: «*Los sistemas basados en la IA pueden consistir simplemente en un programa informático (p. ej. asistentes de voz, programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), pero la IA también puede estar incorporada en dispositivos de hardware (p. ej. robots avanzados, automóviles autónomos, drones o aplicaciones del internet de las cosas)*». Posteriormente, esta definición fue objeto de una revisión técnica por parte del Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, que revisaron y detallaron los criterios para precisar las implicaciones tecnológicas: «*son programas informáticos (y posiblemente también equipos informáticos) diseñados por seres humanos que, dado un objetivo complejo, actúan en la dimensión física o digital mediante la percepción de su entorno mediante la adquisición de datos, la interpretación de los datos estructurados o no estructurados, el razonamiento sobre*

el conocimiento o el tratamiento de la información, fruto de estos datos y la decisión de las mejores acciones que se llevarán a cabo para alcanzar el objetivo fijado». No obstante, en su Libro Blanco sobre la IA de 2020 (*White Paper*, COM(2020) 65 final), la Comisión evitó ofrecer una definición explícita del concepto de «sistema de IA», que retomó más tarde en su propuesta de reglamento.

A modo de reflexión, y aunque se trate de un aspecto tangencial, considero oportuno señalar que, desde una perspectiva económica, el interés de los desarrolladores en que sus sistemas de IA se definan exclusivamente como «software» puede ser complejo, dado que la protección jurídica del software recae en el ámbito del Derecho de autor (art. 10 del Acuerdo sobre los ADPIC), y los algoritmos que contiene, y que constituyen su componente funcional clave y representan su principal valor económico (Gonzales, 2023; Qin et al., 2024), al ser considerados *métodos matemáticos*, no quedan protegidos (artículo 9.2 ADPIC; Foss-Solbrekk, 2021). No obstante, la definición como «software» proporciona mayor certeza jurídica en lo que respecta a la titularidad y gestión de los derechos de propiedad intelectual, y se podría recurrir a otros mecanismos de protección indirecta de los algoritmos, como a través de la protección de secretos empresariales o, en su caso, tal vez, mediante la inclusión de cláusulas contractuales que restrinjan el uso por parte de los licenciarios o usuarios finales de los resultados generados por dichos sistemas. Aunque, en este último caso, dado que los resultados generados por los sistemas de IA no pueden ser predichos ni por el desarrollador ni por el usuario (por todos, Militsyna, 2023), dichos resultados podrían considerarse, en mi opinión, de *dominio público*, lo que implicaría que dichas cláusulas estarían configurando un monopolio *de facto* sobre los algoritmos, algo que contraviene el criterio del legislador ya que, por su naturaleza, no son susceptibles de monopolización legal.

Siguiendo con la definición propuesta por la Comisión, se establece que, mediante «una o varias» de las técnicas y estrategias descritas en el Anexo I, el software puede generar contenidos, predicciones, recomendaciones o decisiones que influyen en el entorno con el que interactúa. Tal y como está formulada esta definición, podría aplicarse también a ciertos tipos de software tradicional. Por ejemplo, si un software genera predicciones o toma decisiones basadas en reglas preprogramadas que influyen en un entorno, como los conocidos SAP ERP u Oracle NetSuite, podrían encajar en la definición propuesta por la Comisión, aunque no se trate de sistemas de IA, sino de sistemas automatizados convencionales. Esta ambigüedad se debe a que la definición no hace referencia a elementos clave que caracterizan los sistemas de IA, como la interacción dinámica, la respuesta autónoma y proactiva o la adap-

tación a cambios después de su programación inicial (OCDE, 2024; Franceschetti, 2018; Ernst, 2020), como veremos en un momento ulterior. Asimismo, en el referido Anexo I, se incluyen técnicas como las «*estrategias estadísticas, estimación bayesiana, métodos de búsqueda y optimización*». Por lo tanto, un software que emplea estrategias estadísticas, como Stata, que genera predicciones y análisis que pueden influir en el entorno, o un software como OpenBUGS, que genera resultados basados en modelos bayesianos, podría ser erróneamente clasificado como un sistema de IA por el hecho de utilizar «una o varias» técnicas mencionadas en el Anexo.

Por otra parte, la definición no menciona ni sugiere la *autonomía*, un elemento esencial en los sistemas de IA, incluso cuando es mínima o muy limitada (como en la IA débil, también conocida como *narrow AI*), sino que se enfoca en la *automatización*. Aunque estos conceptos serán desarrollados en mayor detalle más adelante, cabe avanzar que la automatización implica la ejecución de tareas siguiendo reglas predefinidas, mientras que la autonomía se refiere al grado en que un sistema puede ejecutar tareas sin intervención humana directa, una vez configurado y dentro de los límites y parámetros establecidos, sin perjuicio de la posibilidad de supervisión humana en distintas etapas de su ciclo de vida (OCDE, 2024). En consecuencia, de la definición se deduce que los sistemas descritos quedan limitados a los «objetivos definidos por seres humanos» y a la ejecución de instrucciones programadas, sin alcanzar el nivel de toma de decisiones autónomas. Pero es relevante señalar que, en el campo de la IA general (AGI) y la denominada superinteligencia, se debate activamente sobre estos conceptos, y desarrollos recientes, como el proyecto *AI Scientist*, han demostrado que, dentro de parámetros predefinidos, los sistemas pueden llevar a cabo experimentos y ajustar su ejecución sin intervención humana directa, particularmente en el ámbito de la investigación científica (Tunç, 2024; Castelvechi, 2024), lo que pone en cuestión los límites conceptuales establecidos en la definición de la Comisión.

En relación con lo hasta aquí expuesto, cabe destacar iniciativas legislativas recientes como, por ejemplo, la «Iniciativa para la Ley Federal que regule la Inteligencia Artificial» en México, presentada en 2024, donde algunos de los problemas conceptuales señalados se reflejan, particularmente en el artículo 4, inciso V. Según esta propuesta, los sistemas de IA se definen como «*aquellos que implican el uso y aprovechamiento de las tecnologías de la información para crear programas informáticos capaces de ejecutar cálculos, operaciones, investigaciones o razonamientos comparables a las que realiza la mente humana*». Esta definición resulta excesivamente amplia, ya que podría abarcar cualquier tipo de software, omitiendo además la referencia al hardware, un componente fundamental en los sistemas

integrados, como he señalado previamente. Otras complejidades conceptuales de esta definición se pueden ilustrar a través de un ejemplo práctico: el sistema ORION (*On-Road Integrated Optimization and Navigation*) de la empresa UPS utiliza *tecnologías de IA* para analizar patrones de tráfico y optimizar rutas de entrega, y aunque los resultados que genera pueden parecer análogos a los que una mente humana podría obtener tras razonar sobre el mismo problema, el proceso a través del cual el sistema alcanza dichos resultados no implica razonamiento estratégico o subjetivo *como el humano*, sino que se basa exclusivamente en algoritmos de optimización preprogramados (Holland et al., 2017). Así, aunque los resultados pueden ser similares en términos de efecto práctico, no se puede afirmar que sean «comparables» en el sentido de *replicar* los «cálculos, operaciones, investigaciones o razonamientos» que «realiza la mente humana». Esto, sin perjuicio de la dificultad de delimitar jurídicamente qué se considera *comparable*, dada la complejidad y naturaleza subjetiva del pensamiento humano. Además, el uso de «comparable» en la definición, sin especificar elementos clave que distinguen a los sistemas de IA, como el aprendizaje automático o la toma de decisiones autónoma (OCDE, 2024; Ernst, 2020), abre la posibilidad de que, bajo la denominación de «IA», se incluya incluso software convencional como Microsoft Excel, que realiza cálculos complejos que una «mente humana» podría efectuar, pero lo hace exclusivamente mediante procesos estrictamente automatizados y preprogramados, sin ningún tipo de aprendizaje o decisión autónoma (Saunders, 2018), lo que, en definitiva, debería impedir su clasificación como sistema de IA.

Otro ejemplo relevante es el artículo 4, apartado I, de las *Dispõe sobre o uso da Inteligência Artificial* del Brasil (proyecto de ley núm. 2338/2023, SF/23833.90768-16), que define un sistema de IA como el «*sistema computacional, com graus diferentes de autonomia, desenhado para inferir como atingir um dado conjunto de objetivos, utilizando abordagens baseadas em aprendizagem de máquina e/ou lógica e representação do conhecimento, por meio de dados de entrada provenientes de máquinas ou humanos, com o objetivo de produzir previsões, recomendações ou decisões que possam influenciar o ambiente virtual ou real*». Aunque la definición se refiere a un *sistema computacional*, que en sentido técnico puede incluir tanto software como hardware, parece centrarse casi exclusivamente en los procesos que ejecuta el software, ya que operaciones como la inferencia, predicción y generación de resultados son típicamente realizadas por algoritmos implementados en programas específicos (Sanders, 2019). La mención a que los *datos de entrada* pueden provenir de *máquinas* sugiere una posible interacción con componentes de *hardware*, como sensores

o interfaces físicas. Sin embargo, el núcleo de la definición, como he señalado, se enfoca principalmente, en cómo el software procesa esos datos para generar salidas (predicciones, recomendaciones o decisiones) priorizando el tratamiento algorítmico sobre la interacción directa con el hardware, lo que posiblemente podría limitar la aplicación de la definición a sistemas de IA integrados, como los que se encuentran en robótica o dispositivos industriales, que requieren un enfoque más holístico en el que tanto el software como el hardware sean tratados como componentes inseparables del sistema (Wang et al., 2022; Arents et al., 2022; Tsolakis et al., 2023).

Por su interés, y aunque se trata de unas directrices no vinculantes que fueron elaboradas por la *Conference toward AI Network Society* en agosto de 2019, bajo la supervisión del Ministerio de Asuntos Internos y Comunicaciones de Japón, es pertinente mencionar la definición incluida en estas «*AI Utilization Guidelines*», según las que «*“AI” refers to a concept that collectively refers to AI software and AI systems. “AI software” refers to software that has the function to change its own implementation or output in the process of utilization, by learning from data, information, and knowledge; or by other methods. For example, machine learning software is classified into this category. “AI system” refers to systems that incorporate AI software as a component. For example, robots and cloud systems that implement AI software are classified under this category*» (Conference toward AI Network Society, 2019). La definición presenta un enfoque técnico preciso al destacar que el software es un elemento, pero no el único, en un sistema de IA más amplio, que también incluye hardware y otros componentes necesarios para interactuar con entornos físicos o virtuales, lo que refleja, a mi entender acertadamente, la complejidad estructural que los marcos regulatorios deben tener en cuenta para asegurar que la normativa abarque todas las dimensiones operativas (Weissinger, 2022).

3. EL ROL DE LA AUTONOMÍA Y LAS TÉCNICAS AVANZADAS EN LA DEFINICIÓN JURÍDICA DE LOS SISTEMAS DE IA

La nota interna de 25 de noviembre de 2022 (14954/22 LIMITE), remitida al Comité de Representantes Permanentes (COREPER) del Consejo de la Unión Europea, que precede a la propuesta formal del Consejo sobre el Reglamento de Inteligencia Artificial, hace referencia a un compromiso asumido en el marco de las negociaciones interinstitucionales específicamente orientado, según indica, a «mejorar» la definición de «sistema de IA». Este compromiso se enfoca en restringir su ámbito a aquellos sistemas desarrollados mediante enfoques basados en aprendizaje automático (*ma-*

chine learning) o en *métodos lógicos o basados en el conocimiento*, excluyendo así definitivamente de su alcance los sistemas de software convencionales que, si bien pueden presentar cierta complejidad operativa, no reúnen las características técnicas avanzadas que permiten ser considerados sistemas de IA (*«In order to ensure that the definition of an AI system provides sufficiently clear criteria for distinguishing AI from more classical software systems»*). En este contexto, la propuesta del Consejo de la Unión Europea, de 6 de diciembre de 2022 (ST 15698/22 INIT), define el «sistema de IA» como «un sistema concebido para funcionar con elementos de autonomía que, a partir de datos e información generados por máquinas o por seres humanos, infiere la manera de alcanzar una serie de objetivos, utilizando para ello estrategias de aprendizaje automático o estrategias basadas en la lógica y el conocimiento, y produce información de salida generada por el sistema, como contenidos (sistemas de inteligencia artificial generativa), predicciones, recomendaciones o decisiones, que influyen en los entornos con los que interactúa el sistema de IA». Dicha definición fue incorporada sin cambios relevantes en normativas todavía hoy vigentes como en el artículo 3.3 del Real Decreto español 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (BOE núm. 268, de 9 de noviembre de 2023).

El cambio terminológico de «software» a «sistema» amplía significativamente el alcance de la definición, al permitir incluir tanto el hardware como los algoritmos y los datos que interactúan en el funcionamiento de los sistemas. No obstante, describir un sistema de IA como «sistema» puede generar una impresión de tautología y reflejar una falta de precisión conceptual. Pero es cierto que la definición aporta precisión en aspectos técnicos clave, como el reconocimiento de la *autonomía* como una característica fundamental, lo que constituye, en mi opinión, un avance notable respecto a la propuesta original de la Comisión Europea. La *autonomía*, en este contexto, y aunque la definición no menciona explícitamente *grados o niveles* de autonomía operativa, se refiere al nivel en que un sistema de IA puede ejecutar procesos, realizar ajustes en su funcionamiento y generar resultados sin intervención humana continua, dentro de los parámetros y configuraciones predefinidos por sus desarrolladores (Wischmeyer et al., 2020; OCDE, 2024). La ausencia de intervención humana continua no implica una falta total de supervisión o control humano, sino más bien una disminución en la frecuencia y el grado de intervención directa. Esta conceptualización contrasta con los sistemas *automatizados* convencionales, que típicamente requieren reprogramación o intervención humana constante para

modificar su operatividad (Formosa, 2021). Por ejemplo, un sistema *automatizado* en una línea de montaje, que sigue un conjunto de reglas para ensamblar piezas de forma repetitiva, no puede ajustar su comportamiento si cambia el diseño de la pieza sin ser reprogramado (Jia et al., 2023). Por tanto, aunque en ciertos contextos la distinción entre automatización y autonomía puede parecer sutil, un sistema de IA requiere algún grado de autonomía operativa, aunque sea mínima o muy limitada (*narrow AI*), para ser considerado como tal, pues es lo que lo diferencia de un sistema meramente automatizado (OCDE, 2024). En cualquier caso, esta *independencia operativa* se deriva, en gran medida, de las *técnicas avanzadas* utilizadas en su diseño y funcionamiento (por ejemplo, el aprendizaje automático), y de un mínimo grado de *adaptabilidad*, concepto que se abordará en detalle más adelante. Pero cabe señalar que las *técnicas avanzadas* mencionadas son métodos para procesar datos, generar *modelos* (como los predictivos que utiliza Amazon en su sistema de recomendaciones, por ejemplo) y realizar tareas específicas, pero su uso no necesariamente implica la existencia de *autonomía operativa*. Por ejemplo, un sistema de recomendaciones basado en aprendizaje automático puede generar *predicciones* sin intervención humana directa, pero si cada ajuste o cambio en el algoritmo requiere intervención humana (es decir, recalibración manual), no puede considerarse autónomo (Baiocco et al., 2022). Por ello, en mi opinión, considero que la inclusión explícita del concepto de autonomía en una definición legal rigurosa resulta necesaria.

También en relación con las mencionadas *técnicas avanzadas*, la deliberada restricción de la definición del Consejo a «estrategias de aprendizaje automático» y «estrategias basadas en la lógica y el conocimiento», aunque predominantes en la actualidad (Tufail et al., 2023), establece un marco conceptual que, tal vez, podría no ser lo suficientemente amplio para incorporar tecnologías emergentes que no encajen estrictamente en estas categorías, como los posibles desarrollos en sistemas neuromórficos, por ejemplo (Cutsuridis, 2024; Ivanov, 2022). Tal limitación la encontramos también en otros textos normativos como, por ejemplo, en la antemencionada *Dispõe sobre o uso da Inteligência Artificial* de Brasil, que hace explícita referencia a grados de autonomía («*com graus diferentes de autonomia*»), lo que proporciona una mayor especificidad que la definición del Consejo en cuanto a las variaciones operativas de los sistemas, pero presenta la misma limitación técnica al enfocarse únicamente en «*abordagens baseadas em aprendizagem de máquina e/ou lógica e representação do conhecimento*». En contraste, la Ley de Inteligencia Artificial y Datos de Canadá (*Artificial Intelligence and Data Act*, Bill C-27, 1st Sess., 44th Parl. Can. 2022), que forma parte de la *Digital Charter Implementation Act* de 2022, define un sistema de IA como

«a technological system that, autonomously or partly autonomously, processes data related to human activities through the use of a genetic algorithm, a neural network, machine learning or another technique in order to generate content or make decisions, recommendations or predictions». Los niveles de autonomía se reconocen aquí como un rasgo distintivo («autonomously or partly autonomously»), y la norma ofrece un mayor nivel de especificidad al identificar técnicas concretas (como redes neuronales y algoritmos genéticos) pero dejando abierta de manera explícita la posibilidad de incluir otras metodologías futuras («or another technique»), lo que le confiere una mayor flexibilidad ante tecnologías emergentes.

Asimismo, la propuesta de ley *Safe and Secure Innovation for Frontier Artificial Intelligence Models Act* de California de 2024 (S.B. 1047, 2024 Leg., Reg. Sess. Cal. 2024), también destaca el *nivel de autonomía* como una característica definitoria de los sistemas de IA, pero opta por no mencionar técnicas específicas empleadas en su desarrollo y funcionamiento: «an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments». Resulta llamativo, sin embargo, que otra propuesta del mismo estado, la AB-331 (*Automated Decision Tools Act*, A.B. 331, 2024 Leg., Reg. Sess. Cal. 2024), que regula el uso de herramientas de toma de decisiones automatizadas (ADT, por sus siglas en inglés), ni menciona la autonomía ni identifica técnicas avanzadas específicas en su definición de sistema de IA («means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing a real or virtual environment»), sin introducir elementos distintivos que los diferencien de los sistemas automatizados convencionales.

Otro ejemplo adicional que ilustra las divergencias entre marcos normativos es la *National Defense Authorization Act for Fiscal Year 2024* (NDAA 2024) de Estados Unidos, que adopta la definición de sistema de IA establecida en la *National Artificial Intelligence Initiative Act of 2020* (Pub. L. No. 116-283, Div. E, Tit. I (2021)–NAIIA 2020): «a machine-based system that can for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to: a. Perceive real and virtual environments. b. Abstract such perceptions into models through analysis in an automated manner. c. Use model inference to formulate options for information or action». Esta definición, que muestra una gran similitud con la contenida en el *UK Statutory Instrument 2021 núm. 1264, Schedule 3*, del Reino Unido, no menciona explícitamente la *autonomía operativa* del sistema que le permita funcionar sin

intervención humana. El concepto de «abstraer tales percepciones en modelos mediante análisis automatizado» parece referirse a cómo el sistema transforma los datos del entorno (percepciones) en representaciones simplificadas como modelos matemáticos o *estructuras* abstractas (por ejemplo, redes neuronales o árboles de decisión), lo que sugiere un alto grado de *automatización* dentro de los límites establecidos por humanos, pero ello no implica necesariamente autonomía operativa. Además, la definición enfatiza que los objetivos del sistema son siempre establecidos por seres humanos («*for a given set of human-defined objectives*») excluyendo, al igual que la definición propuesta por la Comisión Europea, la posibilidad de que el sistema ajuste sus objetivos o estrategias de manera *autónoma* ante nuevas circunstancias.

Un caso similar se presenta con la norma técnica ISO/IEC 22989:2023, ratificada por la Asociación Española de Normalización en agosto de 2023, y que aborda conceptos y terminología de inteligencia artificial. Aunque no es vinculante, su inclusión en este estudio creo que es relevante ya que contribuye a la armonización de conceptos en el ámbito técnico internacional. La norma define un sistema de IA como «*engineered system that generates outputs such as content, forecasts, recommendations, or decisions for a given set of human-defined objectives*», y añade que «*the engineered system can use various techniques and approaches related to artificial intelligence to develop a model to represent data, knowledge, processes, etc., which can be used to conduct tasks*». Se trata de una definición deliberadamente amplia y flexible que menciona la posibilidad de que el sistema utilice una variedad de técnicas y enfoques relacionados con la IA sin especificar cuáles en particular. Sin embargo, esta flexibilidad no presupone necesariamente la existencia de *autonomía operativa* pues, como he indicado anteriormente, técnicas avanzadas como el aprendizaje automático o las redes neuronales pueden dotar al sistema de un grado significativo de autonomía, pero su uso no implica por sí mismo la existencia de tal autonomía. En otras palabras, el uso de técnicas avanzadas no convierte automáticamente a un sistema en autónomo, ya que estas pueden emplearse en sistemas que siguen siendo puramente automatizados, es decir, dichas *técnicas avanzadas* pueden ser un medio para lograr esa autonomía, pero no son *sinónimo* de la misma. En este caso, pues, la definición no incluye ni explícita ni implícitamente la autonomía operativa, un elemento que debería ser clave en la definición de los sistemas de IA, por lo expuesto hasta aquí. Además, al igual que en la *National Artificial Intelligence Initiative Act of 2020* y la definición propuesta por el Consejo de la Unión Europea, hace hincapié en que, aunque el sistema puede operar de forma *automatizada*, lo hace dentro de un marco predefinido por humanos («*for a given set of human-defined objectives*»),

excluyendo cualquier sistema que pueda definir sus propios objetivos o estrategias sin intervención humana, lo que ya apunta la necesidad de futuras adaptaciones normativas si se pretende abordar los riesgos emergentes asociados a estos sistemas.

4. INEXACTITUDES Y AMBIGÜEDADES EN LA TRADUCCIÓN DE LAS DEFINICIONES DE SISTEMAS DE IA

El 14 de junio de 2023, el Parlamento Europeo adoptó enmiendas a la propuesta de Reglamento sobre Inteligencia Artificial de la Comisión Europea, recogidas en el documento P9_TA(2023)0271, que proponen una modificación de la definición de sistema de IA, en los siguientes términos: *«un sistema basado en máquinas diseñado para funcionar con diversos niveles de autonomía y capaz, para objetivos explícitos o implícitos, de generar información de salida—como predicciones, recomendaciones o decisiones—que influya en entornos reales o virtuales»*. Esta definición mantiene mayor coherencia con las definiciones de la OCDE (2019) y la contenida en el Documento de Trabajo del Personal de la Comisión (Impact Assessment, SWD(2021) 84 final) que la propia definición propuesta por la Comisión.

En la versión original en inglés se utiliza la expresión «machine-based system», que fue traducida al español como «sistema basado en máquinas», lo que presenta, en mi opinión, problemas de precisión técnica y jurídica. En inglés, el término «machine» según el diccionario Merriam-Webster, no se limita a dispositivos físicos o mecánicos sino incluye sistemas electrónicos y computacionales, e incluso estructuras organizativas o combinaciones de partes que interactúan para realizar tareas específicas (Merriam-Webster, 2024), por lo que esta definición es pertinente en el contexto de los sistemas de IA, que integran hardware, software y algoritmos. En contraste, la Real Academia Española (RAE) define «máquina» en términos *predominantemente mecánicos*, limitándola a dispositivos que transforman energía o a conjuntos de partes físicas (RAE, 2014), lo que hace que la traducción «sistema basado en máquinas» sea imprecisa para describir la naturaleza de los sistemas de IA. Además, el uso del plural «máquinas» en la traducción al español podría generar interpretaciones erróneas al sugerir la existencia de múltiples dispositivos independientes, cuando en el contexto de la IA nos referimos generalmente a sistemas integrados donde las distintas partes interactúan de manera coordinada para cumplir con los «objetivos explícitos o implícitos» mencionados en la definición (Brachman et al., 2020; OCDE, 2024). Como analizaremos en el siguiente apartado, el texto del Reglamento finalmente aprobado suprime el plural y se refiere a «una

máquina», lo cual también presenta ciertas problemáticas. En todo caso, esta posible falta de precisión en la traducción al español de la definición del Parlamento no se compensa, a mi entender, en el resto de la definición, como analizaremos seguidamente. Y del mismo modo que la traducción de *machine learning* es *aprendizaje automático*, evitando así la confusión potencial derivada de una traducción literal, considero que *machine-based system* no debería traducirse literalmente, sino que debería emplearse una terminología que refleje con mayor exactitud la naturaleza compleja e integrada de los sistemas de IA, lo será objeto de un análisis detallado al abordar la definición establecida en el texto definitivo del Reglamento europeo.

A diferencia de la definición propuesta por el Consejo de la UE, la definición del Parlamento no solo menciona la autonomía de los sistemas, sino que introduce también la existencia de *niveles o grados de autonomía*, y como he apuntado anteriormente, es un matiz relevante porque reconoce que la independencia de los sistemas de IA en relación al control humano puede situarse en un espectro, lo que permite un enfoque más preciso para evaluar sus riesgos y responsabilidades. Además, la definición evita referirse a técnicas específicas, como el aprendizaje automático o estrategias basadas en la lógica y el conocimiento, optando por un enfoque *más abstracto* que pretende abarcar un rango amplio de tecnologías de IA, una abstracción que, sin embargo, no compromete la distinción entre sistemas de IA y sistemas automatizados convencionales puesto que existe la referencia explícita a los diferentes niveles de autonomía. Todo ello contrasta con la postura del Consejo de la UE que, como analizamos anteriormente, restringió deliberadamente la definición a sistemas basados en las mencionadas técnicas, sin referencia alguna a la existencia de *autonomía operativa*.

Por otro lado, la definición del Parlamento, a diferencia de las del Consejo y de la Comisión, introduce la referencia a los *objetivos implícitos* («objetivos explícitos e implícitos»), que alude a aquellos objetivos que, aunque no estén formulados de manera directa, pueden inferirse a partir de las instrucciones, los datos o el contexto en el que opera el sistema (Masters et al., 2021). Así, un sistema que sigue reglas proporcionadas por sus desarrolladores puede alcanzar objetivos implícitos que, aunque no están programados directamente, se deducen de cómo el sistema aplica esas reglas en su entorno operativo (OCDE, 2024). Del mismo modo, en sistemas *entrenados* con grandes conjuntos de datos, los objetivos pueden inferirse a partir de los patrones observados en esos datos. Y, por ejemplo, en el caso de los sistemas que utilizan técnicas de *aprendizaje por refuerzo*, es decir, algoritmos que permiten a un *agente* (generalmente un algoritmo) ajustar su comportamiento mediante recompensas o penalizaciones (AlMahamid et al., 2022), los objetivos específicos

no siempre están completamente definidos desde el principio, sino que se ajustan conforme el sistema interactúa con su entorno o con los usuarios (OCDE, 2024). No obstante, esta flexibilidad operativa no implica que los sistemas puedan determinar sus propios objetivos o fines de forma autónoma, ya que siempre están limitados por los parámetros fijados por los desarrolladores. Por consiguiente, cualquier sistema que pudiera establecer sus propios objetivos, en el sentido de poder fijar metas sin intervención humana, quedaría fuera del alcance de la definición propuesta por el Parlamento. También en el caso del *UK Statutory Instrument 2021 núm. 1264, Schedule 3*, del Reino Unido, que establece que los sistemas de IA pueden hacer «recommendations, predictions or decisions; with a view to achieving a specific objective» sin especificar si dichos objetivos son explícitos o implícitos, o si son o no objetivos exclusivamente definidos por seres humanos, pero su referencia inicial a la «programación o entrenamiento» del sistema («*Technology enabling the programming or training of a device or software to*») comporta que los objetivos dependan necesariamente de la configuración realizada por los desarrolladores humanos. Por tanto, podemos concluir que, hasta la fecha, la autonomía reconocida en las definiciones que la contemplan ya sea de manera explícita o implícita, sigue estando estrictamente subordinada al control humano (Tunç, 2024).

Continuando con el análisis de la definición del Parlamento Europeo, la versión original en inglés menciona que el sistema de IA podría influir tanto en los entornos *físicos* como en los *virtuales* («physical or virtual»). El *entorno*, ya sea físico como el mundo material, o virtual como los espacios digitales, es aquel en el que el sistema recibe datos y actúa (ejerce *influencia*, según la definición del Parlamento), ya sea mediante dispositivos físicos o realizando acciones automatizadas a partir de datos (OCDE, 2024), como patrones de comportamiento o señales biológicas, como pueden ser, por ejemplo, el ritmo cardíaco o la actividad cerebral (Bhardwaj et al., 2022; Li et al., 2024). Sin embargo, la traducción de la definición del Parlamento al español emplea el término «reales» en lugar de «físicos» («reales o virtuales»), lo que introduce una dicotomía entre lo «real» y lo «virtual» inadecuada desde un punto de vista técnico y jurídico. Los entornos virtuales también son *reales* en la medida en que generan entradas y permiten acciones operativas; en otras palabras, tanto los entornos físicos como los virtuales forman parte de la realidad operativa del sistema de IA.

Otro problema en términos de precisión lingüística y jurídica de la versión en español, es la introducción del término «capaz» («capaz, para objetivos explícitos o implícitos, de generar información»), que difiere de la formulación original en inglés, donde se utiliza la expresión «*that can*», más neutral y técnica y que se re-

fiere únicamente a la posibilidad de que el sistema realice determinadas acciones, sin atribuirle una capacidad inherente. Traducir «that can» como «que puede» mantendría, en mi opinión, aquella neutralidad, asegurando una mayor precisión y evitando malentendidos sobre la naturaleza operativa del sistema.

Por último, al referirse a los resultados generados por los sistemas, tanto la definición del Parlamento como la de la Comisión Europea establecen una lista *ejemplificativa* («como predicciones, recomendaciones o decisiones»), mientras que la propuesta de la Comisión, además, incluye expresamente la generación de *contenidos*, ampliando su alcance a *sistemas de IA generativa*, como lo son ChatGPT (OpenAI) o Midjourney. Esta inclusión fue corroborada por el Consejo de la UE, que señaló en su propuesta de definición que un sistema de IA «*produce información de salida generada por el sistema, como contenidos (sistemas de inteligencia artificial generativa)*». De igual forma, la última versión de la definición propuesta por la OCDE y la norma técnica ISO/IEC 22989:2023 presentan una lista abierta («*such as [...]*»), añadiendo una mención específica a la generación de *contenidos*. En contraste, la *Artificial Intelligence and Data Act* de Canadá menciona la generación de contenidos, pero sin sugerir listas abiertas o ejemplificativas («*in order to generate content or make decisions, recommendations or predictions*»). Y otras normativas, como la *Dispõe sobre o uso da Inteligência Artificial* de Brasil («*com o objetivo de produzir previsões, recomendações ou decisões*»), el *UK Statutory Instrument 2021 núm. 1264, Schedule 3* del Reino Unido («*make recommendations, predictions or decisions*»), o la *National Artificial Intelligence Initiative Act of 2020* de Estados Unidos («*make predictions, recommendations or decisions*»), no sugieren ni listas abiertas ni mencionan la generación de «contenido», lo que, interpretado de manera restrictiva, podría limitar su aplicación a sistemas que no generen resultados más allá de los expresamente citados (predicciones, recomendaciones y decisiones). Ahora bien, cuando los sistemas de IA generativa producen *contenido*, es decir, productos tangibles como texto, imágenes, audio, vídeo o, por ejemplo, código informático, que pueden ser utilizados directamente por humanos o integrados en sistemas automatizados (Gondwe, 2023; Kalota, 2024; Banh et al., 2023), este proceso se basa, en un nivel fundamental, esencialmente en el uso de algoritmos que incluyen predicciones, recomendaciones y toma de decisiones como componentes esenciales de su funcionamiento. En tal sentido, una *predicción* es la anticipación de un resultado basada en patrones y datos procesados por el sistema, sin que ello implique una acción autónoma por parte del sistema (OCDE, 2024), y en el caso de los sistemas generativos de texto como en los basados en modelos probabilísticos (por ejemplo, GPT) el proceso se basa en la *predicción* de la palabra más probable que sigue en una

secuencia, utilizando el contexto de las palabras precedentes. En la generación de imágenes, el sistema anticipa el píxel más probable basándose en los patrones previamente *aprendidos* a partir de conjuntos de datos, y este mecanismo se extiende a otros tipos de contenido, como notas musicales, elementos visuales o cualquier otra unidad de información (Abood et al., 2024). La *recomendación*, por su parte, comporta que el sistema sugiera un curso de acción o una solución en función de los datos procesados (OCDE, 2024), lo que en los sistemas generativos puede manifestarse cuando, por ejemplo, se genera un texto sobre un tema específico y sugiere ciertos enfoques o perspectivas basándose en los patrones *aprendidos* o, en la generación de imágenes, sugiere estilos o elementos visuales que se ajustan a la descripción proporcionada o al contexto establecido. Finalmente, la toma de *decisiones* se refiere a que el sistema actúa directamente sobre su entorno (OCDE, 2024), lo que en el contexto de la generación de contenido puede traducirse en decidir qué palabras, frases, estructuras, composiciones, colores, notas musicales o, por ejemplo, efectos visuales utilizar, guiado por las predicciones y los parámetros definidos por los desarrolladores.

En consecuencia, considero que los sistemas generativos deben contemplarse dentro de la definición de sistema de IA incluso cuando no se haga una referencia explícita a la generación de *contenido*. Sin embargo, lo contrario no es aplicable: no todos los sistemas de IA que realizan predicciones, recomendaciones o decisiones generan *contenido* en los términos antemencionados. Por ejemplo, un sistema de predicción meteorológica como el del *European Centre for Medium-Range Weather Forecasts* puede generar una *predicción*, es decir, un dato numérico o estimación, que no necesariamente constituye un «contenido» en el sentido que lo hacen los sistemas de IA generativa (Chantry et al., 2023). En mi opinión, la razón por la que se menciona la generación de *contenido* en las definiciones legales y técnicas antemencionadas tiene como objetivo destacar la relevancia de los sistemas generativos diferenciándolos de otros sistemas que, aunque también se basan en predicciones, recomendaciones o decisiones, no producen *contenido* de la misma naturaleza ni con la misma finalidad.

5. ANÁLISIS ESTRUCTURAL DE LOS SISTEMAS: MODELOS, PROCESOS DE INFERENCIA Y TERMINOLOGÍAS TÉCNICAS

La definición establecida en el texto definitivo del Reglamento (UE) 2024/1689, adoptado el 13 de junio de 2024, dispone que un sistema de IA es «*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 objeti-*

vos 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». En la versión original en inglés se mantiene la expresión «machine-based system», como en la definición propuesta por el Parlamento Europeo, sin embargo, la nueva traducción al español introduce una variación al cambiar el plural «máquinas» al singular, refiriéndose ahora a «un sistema basado en una máquina». Como he argumentado previamente, el uso de «máquina», tanto en singular como en plural, resulta problemático y, en cierto modo, incluso redundante, en el sentido de que podría afirmarse cualquier sistema de IA requiere, por su propia naturaleza, algún tipo de infraestructura tecnológica subyacente. El uso del singular podría inducir, erróneamente, a pensar que el sistema se reduce a un único dispositivo físico, cuando en realidad muchos sistemas modernos operan en entornos distribuidos que combinan infraestructuras como la *nube*, redes locales, y otros componentes y, en estos entornos, el procesamiento de datos y la ejecución de algoritmos se realizan en plataformas distribuidas, que incluyen servidores, hardware especializado y redes de comunicación (Mehmood, 202 et al.; Mungoli, 2023). Cabe mencionar que el Considerando 12 del reglamento afirma que «el término “basado en una máquina” se refiere al hecho de que los sistemas de IA se ejecutan en máquinas», parece que tratando de corregir la posible impresión de que están limitados a un único dispositivo. Aquí, el uso del plural «máquinas» en el Considerando entra en contradicción con el uso del singular en la definición principal (art. 3), que, recordemos, modifica el plural («máquinas») utilizado en la traducción al español de la definición de la propuesta del Parlamento (art. 3), lo cual contribuye a generar confusión. Los considerandos pueden proporcionar contexto y orientación cuando el texto normativo es ambiguo, pero no pueden contradecir las disposiciones del reglamento (UE, 2015).

La expresión «machine-based system», aunque utilizada previamente en los *Principios sobre IA* de la OCDE (2019) y la *National Artificial Intelligence Initiative Act* de EE. UU. de 2020, no representa la terminología más precisa ni extendida en el contexto actual. Por ejemplo, la norma ISO/IEC 22989:2023 se refiere a «engineered system», que abarca cualquier sistema diseñado deliberadamente bajo principios científicos y técnicos, optimizado conforme a principios de *ingeniería* (Bajzek et al., 2021). En cambio, la *Artificial Intelligence and Data Act* de Canadá introduce el concepto de «technological system», más general, que reconoce la complejidad de la infraestructura técnica de estos sistemas. En contraste, la *Dispõe sobre o uso da Inteligência Artificial* de Brasil utiliza la expresión «sistema computacional»,

poniendo un énfasis quizá restrictivo en el procesamiento algorítmico. Por otra parte, la *Safe and Secure Innovation for Frontier Artificial Intelligence Models Act* de California de 2024, combina términos en la expresión «*engineered or machine-based system*», lo que crea, en mi opinión, una falsa dicotomía técnica y conceptual, pues todo sistema de IA diseñado conforme a *principios de ingeniería* es, *de facto*, *machine-based*, ya que depende de una infraestructura física (servidores, hardware especializado, o redes distribuidas), y todo *machine-based system* con capacidades avanzadas (por ejemplo, procesamiento de grandes volúmenes de datos, la ejecución de algoritmos de aprendizaje profundo o la interacción en tiempo real con múltiples fuentes de información), requiere de un diseño técnico estructurado, lo que también lo convierte en un *engineered system*. Así, la conjunción «or» en lugar de aportar claridad, genera confusión al separar dos conceptos que son, en la práctica, interdependientes. Tal vez, por lo dicho, una formulación más coherente y precisa sería la adopción exclusiva de «*engineered system*», que integra tanto el diseño técnico como la infraestructura física necesaria para su funcionamiento.

Por otra parte, la definición del Reglamento europeo, además de referirse explícitamente a los «distintos niveles de autonomía» del sistema, introduce el concepto de *adaptación* del sistema una vez implementado en un entorno operativo real (despliegue) para cumplir con sus objetivos. Esta adaptación permite que el sistema ajuste automáticamente los parámetros de su *modelo matemático* o *de aprendizaje* (como los utilizados en AlphaGo de DeepMind, por ejemplo), es decir, el conjunto de reglas y algoritmos que utiliza para procesar datos y extraer patrones, mediante el análisis continuo de nuevas entradas de datos durante su entrenamiento (Koltsakis et al., 2023; Floridi et al., 2020), con el fin de optimizar el cumplimiento de los objetivos predefinidos por los diseñadores (Fischer, 2023; OCDE, 2024). Esta *adaptación* no implica que el sistema establezca nuevos objetivos de manera autónoma en los términos que anteriormente expuse, sino que los ajustes se realizan dentro de los límites establecidos, mejorando áreas como la precisión de las *predicciones*, la relevancia de las *recomendaciones* y la toma de *decisiones*. Por ejemplo, un sistema de IA diseñado para la *recomendación de contenido* puede actualizar su modelo en función de los datos adicionales sobre los usuarios, ajustando sus recomendaciones y predicciones para mejorar la relevancia, pero sin alterar los objetivos principales del sistema (OCDE, 2024). El análisis de los referidos modelos lo abordaré nuevamente más adelante al tratar el proceso de inferencia, que es el núcleo de cualquier sistema de IA.

Cabe mencionar que la versión en español del Reglamento, traduce la expresión «that may exhibit adaptiveness» como «que puede mostrar capacidad de

adaptación», lo que introduce una connotación impropia, al sugerir una cualidad inherente al sistema que no se ajusta al enfoque técnico del Reglamento. Una traducción más precisa y neutral creo que sería «que puede adaptarse», evitando así interpretaciones erróneas sobre la naturaleza de los sistemas.

También desde una perspectiva lingüística, la traducción al español ha corregido la inexactitud que aparecía en la versión en español del texto del Parlamento Europeo de 2023, donde se utilizaba el término «reales» en lugar de «físicos» para referirse a los entornos («reales o virtuales»). Como ya señalé anteriormente, esto era problemático porque sugería una distinción entre lo «real» y lo «virtual» cuando, en realidad, el contraste debía hacerse entre entornos «físicos» y «virtuales». La nueva traducción de la expresión «physical and virtual» es «físicos o virtuales», y resuelve aquella confusión conceptual.

Por último, el legislador europeo ha incorporado, acertadamente en mi opinión pues es el núcleo del funcionamiento de estos sistemas, una descripción explícita del *proceso de inferencia*, indicando que el sistema no solo recibe datos, sino que los transforma mediante algoritmos para producir resultados específicos, como predicciones, contenidos, recomendaciones o decisiones («infiere de la información de entrada que recibe la manera de generar resultados de salida [...]'). Ninguna otra definición de sistema de IA de las revisadas en este estudio incluye este nivel de detalle. Ciertamente, describir el proceso de inferencia no es esencial como componente técnico independiente, siendo preferible delimitar el objeto normativo por su funcionalidad (procesar datos, generar salidas) en lugar de cómo lo hace. No obstante, en regulaciones que aborden aspectos relacionados con la transparencia o la responsabilidad, la inferencia debe incluirse por su relevancia para la supervisión de la toma de decisiones automatizadas.

La traducción al español en este apartado es, en general, fiel al texto original en inglés, aunque utiliza la expresión «de la información de entrada» en lugar de «input», un término técnico común que, en mi opinión, podría haberse utilizado para preservar la precisión técnica. Y, asimismo, la traducción de «how to generate outputs» como «la manera de generar resultados de salida», no distorsiona significativamente el sentido y es técnicamente correcto, aunque considero que sería más conciso «la manera de generar salidas», eliminando la redundancia implícita en «resultados de salida».

El proceso de inferencia transforma las entradas en resultados, determinando la opción más adecuada según la información disponible (OCDE, 2024) y, como establece el Considerando 12 del Reglamento, la inferencia no se limita al proce-

samiento de datos, sino que aplica estrategias de aprendizaje automático y lógicas para ajustar los modelos o algoritmos, optimizando los resultados conforme a los objetivos fijados por los desarrolladores. Los datos de entrada pueden proceder de diversas fuentes, como sensores, bases de datos, reglas predefinidas o conocimiento humano, y pueden integrarse tanto en la fase de desarrollo (*entrenamiento* del modelo) como en la fase operativa (después del despliegue), permitiendo la integración continua de nuevas entradas en tiempo real para generar resultados actualizados (OCDE, 2024). En este contexto del proceso de inferencia, un modelo de IA (que hemos definido previamente al tratar la adaptación), es un componente esencial dentro un sistema más amplio que puede incluir dispositivos *de entrada*, como sensores u otros mecanismos que proporcionan datos del entorno físico o virtual, y dispositivos *de salida*, como actuadores que ejecutan acciones físicas o generan respuestas digitales (por ejemplo, mostrar texto en una página web o modificar elementos gráficos en una interfaz), permitiendo así la interacción con el entorno y el procesamiento de datos para generar los resultados deseados (Floridi et al., 2020). Así, por ejemplo, ChatGPT-4 es un modelo de IA que procesa texto, infiere respuestas y genera salidas, y cuando se interactúa con este modelo en una plataforma, ya sea una aplicación o una página web, se está utilizando un sistema de IA que integra dicho modelo con otros componentes, como la interfaz de usuario, los servidores que gestionan las peticiones y el almacenamiento de datos. En este caso, el modelo de IA dentro del sistema realiza las inferencias necesarias para generar las salidas, pero es el sistema de IA en su conjunto el que recoge las entradas, las procesa y actúa sobre los entornos físicos o virtuales.

6. A MODO DE REFLEXIONES FINALES: ¿HACIA UN CONSENSO GLOBAL EN LA REGULACIÓN?

El análisis de las definiciones de los sistemas de IA revela una considerable fragmentación conceptual y terminológica, reflejo de la complejidad inherente a esta tecnología emergente. Aunque parece existir cierto consenso en cuanto a que estos sistemas generan predicciones, recomendaciones y decisiones, persisten discrepancias notables sobre su naturaleza exacta y las tecnologías que deben incluirse en su definición, lo que también se manifiesta en las diferencias normativas en aspectos clave como la automatización y la adaptabilidad, pues mientras algunas normativas omiten referencias explícitas a la automatización, la mayoría no consideran la adaptabilidad ni explicitan el proceso de inferencia.

De acuerdo con lo expuesto en este trabajo, una definición de «sistema de IA» considero que debería incluir, como elementos fundamentales, su *naturaleza tecnológica* como sistema diseñado para procesar información de manera automatizada, ya sea mediante software y, cuando corresponda, hardware integrado; los *niveles de autonomía*, que lo distinguen de los sistemas automatizados convencionales; su *adaptabilidad* a nuevas entradas o variaciones en el entorno sin intervención humana directa; y, posiblemente, el *proceso de inferencia*, mediante el que transforma los datos en resultados específicos utilizando algoritmos de aprendizaje automático, modelos matemáticos u otros métodos de análisis predefinidos.

Tal y como señala el Considerando 12 del Reglamento europeo, es necesaria una definición precisa del concepto de «sistema de IA» para garantizar la seguridad jurídica y facilitar la interoperabilidad global, aunque, esta armonización deba anticipar desarrollos tecnológicos que resultan difíciles de prever con exactitud. Pero esta búsqueda de una definición consensuada a nivel internacional no debe impedir la regulación; al contrario, debe servir como base para establecer el núcleo común que promueva la cooperación internacional, equilibrando la especificidad local con la armonización global, para garantizar un marco regulatorio efectivo y adaptable a largo plazo, y sin obstaculizar la innovación.

7. REFERENCIAS BIBLIOGRÁFICAS

- Abood, M. A., Kadhem, S. M. (2024). A survey of next words prediction models. En J. Rasheed, A. M. Abu-Mahfouz, M. Fahim (Eds.), *Forthcoming networks and sustainability in the AIoT era. Lecture Notes in Networks and Systems* (Vol. 1035, pp. 165-185). Springer. https://doi.org/10.1007/978-3-031-62871-9_14
- AlMahamid, F., Grolinger, K. (2022). Reinforcement learning algorithms: An overview and classification. *arXiv*. <https://doi.org/10.48550/arXiv.2209.14940>. (Trabajo presentado en 2021 IEEE Canadian Conference on Electrical and Computer Engineering, CCECE 2021, pp. 1-7).
- Arents, J., Greitans, M. (2022). Smart industrial robot control trends, challenges and opportunities within manufacturing. *Applied Sciences*, 12(2), 937. <https://doi.org/10.3390/app12020937>
- Baiocco, S., Fernández-Macías, E., Rani, U., Pesole, A. (2022). *The algorithmic management of work and its implications in different contexts* (Documento de referencia No. 9). Proyecto conjunto UE-OIT “Construyendo alianzas sobre el futuro del trabajo” <https://acortar.link/8GESVt>
- Banh, L., Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, 33, 63. <https://doi.org/10.1007/s12525-023-00680-1>

- Bajzek, M., Fritz, J., Hick, H. (2021). Systems engineering principles. En H. Hick, K. Küpper, H. Sorger (Eds.), *Systems engineering for automotive powertrain development* (pp. 149–194). Springer. https://doi.org/10.1007/978-3-319-99629-5_7
- Bhardwaj, A., Kishore, S., Pandey, D. K. (2022). Artificial intelligence in biological sciences. *Life*, 12(9), 1430. <https://doi.org/10.3390/life12091430>
- Brachman, R., Gunning, D., Burke, M. (2020). Integrated artificial intelligence systems: DARPA's impact on AI. *AI Magazine*, 41(2), 66–81. <https://doi.org/10.1609/aimag.v41i2.5300>
- Castelvecchi, D. (2024). Researchers built an 'AI Scientist'—what can it do? *Nature*, 633(266). <https://doi.org/10.1038/d41586-024-02842-3>
- Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., Floridi, L. (2018). Artificial intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics*, 24(2), 505–528. <https://doi.org/10.1007/s11948-017-9901-7>
- Chantry, M., Bouallegue, Z. B., Magnusson, L., Maier-Gerber, M., Dramsch, J. (2023, 20 de junio). The rise of machine learning in weather forecasting. *ECMWF Science Blog*. <https://acortar.link/NiKoo8>
- Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, 6, Article 1421273. <https://doi.org/10.3389/fhumd.2024.1421273>
- Cutsuridis, V. (2024). Neuromorphic cognitive learning systems: The future of artificial intelligence? *Cognitive Computation*, 16(5), 1433–1435. <https://doi.org/10.1007/s12559-024-10308-x>
- Ernst, C. (2020). Artificial intelligence and autonomy: Self-determination in the age of automated systems. En T. Wischmeyer T. Rademacher (Eds.), *Regulating artificial intelligence* (pp. 53–73). Springer. https://doi.org/10.1007/978-3-030-32361-5_3
- Fischer, G. (2023). Adaptive and adaptable systems: Differentiating and integrating AI and EUD. En L. D. Spano, A. Schmidt, C. Santoro, S. Stumpf (Eds.), *End-user development. IS-EUD 2023. Lecture notes in computer science* (Vol. 13917, pp. 3–22). Springer. https://doi.org/10.1007/978-3-031-34433-6_1
- Floridi, L. (2021). The European Legislation on AI: A Brief Analysis of Its Philosophical Approach. *Philosophy & Technology*, 34(2), 215–222. <https://doi.org/10.1007/s13347-021-00460-9>
- Floridi, L., Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Floridi, L., Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds & Machines*, 30(4), 681–694. <https://doi.org/10.1007/s11023-020-09548-1>

- Formosa, P. (2021). Robot autonomy vs. human autonomy: Social robots, artificial intelligence (AI), and the nature of autonomy. *Minds & Machines*, 31(4), 595–616. <https://doi.org/10.1007/s11023-021-09579-2>
- Foss-Solbrekk, K. (2021). Three routes to protecting AI systems and their algorithms under IP law: The good, the bad and the ugly. *Journal of Intellectual Property Law & Practice*, 16(3), 247–258. <https://doi.org/10.1093/jiplp/jpab033>
- Franceschetti, D. R. (Ed.). (2018). *Principles of robotics & artificial intelligence*. Salem Press; Grey House Publishing.
- Gondwe, G. (2023). Exploring the multifaceted nature of generative AI in journalism studies: A typology of scholarly definitions. *SSRN*. <https://doi.org/10.2139/ssrn.4465446>
- Gonzales, J. T. (2023). Implications of AI innovation on economic growth: A panel data study. *Journal of Economic Structures*, 12(13). <https://doi.org/10.1186/s40008-023-00307-w>
- Grieman, K., Early, J. (2023). A risk-based approach to AI regulation: System categorisation and explainable AI practices. *SCRIPTed*, 20(1), 56. <https://script-ed.org/article/a-risk-based-approach-to-ai-regulation-system-categorisation-and-explainable-ai-practices/>
- Holland, C., Levis, J., Nuggehalli, R., Santilli, B., Winters, J. (2017). UPS optimizes delivery routes. *Interfaces*, 47(1), 8–23. <https://doi.org/10.1287/inte.2016.0875>
- Ivanov, D., Chezhegov, A., Kiselev, M., Grunin, A., Larionov, D. (2022). Neuromorphic artificial intelligence systems. *Frontiers in Neuroscience*, 16. <https://doi.org/10.3389/fnins.2022.959626>
- Jia, G., Zhang, Y., Shen, S., Liu, B., Hu, X., Wu, C. (2023). Load balancing of two-sided assembly line based on deep reinforcement learning. *Applied Sciences*, 13(13), 7439. <https://doi.org/10.3390/app13137439>
- Jurafsky, D., Martin, J. H. (2021). *Speech and language processing* (3^a ed.). Stanford University. https://web.stanford.edu/~jurafsky/slp3/ed3bookaug20_2024.pdf
- Kalota, F. (2024). A primer on generative artificial intelligence. *Education Sciences*, 14(2), 172. <https://doi.org/10.3390/educsci14020172>
- Koltsakis, E., Klontzas, M. E., Karantanas, A. H. (2023). What is artificial intelligence: History and basic definitions. En M. E. Klontzas, S. C. Fanni, E. Neri (Eds.), *Introduction to artificial intelligence. Imaging informatics for healthcare professionals* (pp. 1–11). Springer. https://doi.org/10.1007/978-3-031-25928-9_1
- LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Li, A., Li, H., Yuan, G. (2024). Continual learning with deep neural networks in physiological signal data: A survey. *Healthcare*, 12(2), 155. <https://doi.org/10.3390/healthcare12020155>
- OCDE. (2024). *Explanatory memorandum on the updated OECD definition of an AI system*. OECD Artificial Intelligence Papers. <https://oecd.ai/en/work/ai-system-definition-update>

- Masters, P., Vered, M. (2021). What's the context? Implicit and explicit assumptions in model-based goal recognition. En Z.-H. Zhou (Ed.), *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21* (pp. 4516-4523). International Joint Conferences on Artificial Intelligence Organization. <https://doi.org/10.24963/ijcai.2021/615>
- Mehmood, R., Sitek, P., Julián, V., Rashid, M. (2020). Distributed Artificial Intelligence-as-a-Service (DAIaaS) for smarter IoE and 6G environments. *Sensors*, 20(20), 5796. <https://doi.org/10.3390/s20205796>
- Merriam-Webster. (s.f.). *Merriam-Webster.com dictionary*. <https://www.merriam-webster.com/dictionary/machine>
- Militsyna, K. (2023). Human Creative Contribution to AI-Based Output – One Just Can('t) Get Enough. *GRUR International*, 72(10), 939–949. <https://doi.org/10.1093/grurint/ikado75>
- Mirzadeh, I., Alizadeh, K., Shahrokhi, H., Tuzel, O., Bengio, S., Farajtabar, M. (2024). GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models. *arXiv*. <https://doi.org/10.48550/arXiv.2410.05229>
- Mungoli, N. (2023). Scalable, Distributed AI Frameworks: Leveraging Cloud Computing for Enhanced Deep Learning Performance and Efficiency. *arXiv*. <https://doi.org/10.48550/arXiv.2304.13738>
- Pino Diez, R., Gómez Gómez, A., Abajo Martínez, N. (2001). *Introducción a la Inteligencia Artificial: Sistemas Expertos, redes neuronales artificiales y computación evolutiva*. Servicio de Publicaciones de la Universidad de Oviedo.
- Qin, Y., Xu, Z., Wang, X., Skare, M. (2024). Artificial intelligence and economic development: An evolutionary investigation and systematic review. *Journal of the Knowledge Economy*, 15, 1736–1770. <https://doi.org/10.1007/s13132-023-01183-2>
- Radanliev, P., De Roure, D., Van Kleek, M., Walton, J. (2021). Artificial intelligence in cyber physical systems. *AI & Society*, 36(4), 783–796. <https://doi.org/10.1007/s00146-020-01049-0>
- Sanders, N. E. (2019). A balanced perspective on prediction and inference for data science in industry. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.644ef4a4>
- Saunders, M. (2018). Data analysis, machine learning and Microsoft Excel. *MSDN Magazine*, 33(1). <https://learn.microsoft.com/en-us/office/dev/add-ins/excel/excel-add-ins-overview>
- Talib, M. A., Majzoub, S., Nasir, Q., Ghafoor, K. (2021). A systematic literature review on hardware implementation of artificial intelligence algorithms. *The Journal of Supercomputing*, 77(3), 1897–1938. <https://doi.org/10.1007/s11227-020-03325-8>
- Tsolakis, N., Gasteratos, A. (2023). Sensor-driven human-robot synergy: A systems engineering approach. *Sensors*, 23(1), 21. <https://doi.org/10.3390/s23010021>
- Tufail, S., Riggs, H., Tariq, M., Sarwat, A. I. (2023). Advancements and challenges in machine learning: A comprehensive review of models, libraries, applications, and algorithms. *Electronics*, 12(8), 1789. <https://doi.org/10.3390/electronics12081789>

- Tunç, A. (2024). Can AI determine its own future? *AI & Society*. <https://doi.org/10.1007/s00146-024-01892-5>
- Wallach, W., Marchant, G. E. (2018). An agile ethical/legal model for the international and national governance of AI and robotics. *Association for the Advancement of Artificial Intelligence*. https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_77.pdf
- Wang, J., Herath, D. (2022). What makes robots? Sensors, actuators, and algorithms. En D. Herath, D. St-Onge (Eds.), *Foundations of robotics*. Springer. https://doi.org/10.1007/978-981-19-1983-1_7
- Weissinger, Laurin B. (2024). *AI, Complexity, and Regulation*. En Justin B. Bullock et al. (Eds.), *The Oxford Handbook of AI Governance* (Oxford Handbooks). Oxford Academic. <https://doi.org/10.1093/oxfordhb/9780197579329.013.66>
- Wischmeyer, T., Rademacher, T. (Eds.). (2020). *Regulating Artificial Intelligence*. Springer Nature. <https://doi.org/10.1007/978-3-030-32361-5>