INTRODUCCIÓN*

El objeto de estudio de este trabajo aglutina dos realidades que hasta hace poco tiempo se han movido en dimensiones muy diferentes y distantes, pero en la actual sociedad algorítmica se ven obligadas a interactuar de manera cada vez más intensa: la inteligencia artificial (en adelante IA) y el Derecho Constitucional.

Esta interacción supone tres dificultades añadidas a todo análisis de Derecho Constitucional clásico. En primer lugar, debemos familiarizarnos con los elementos más esenciales de una disciplina ajena a todo jurista como es la IA. El problema de la definición misma de la IA da cuenta del reto que supone[1]. En segundo lugar, la anterior tarea se complica con la exponencial evolución tecnológica, cuantitativa, pero sobre todo cualitativa de los últimos años en torno a la IA. El constante cambiante telón de fondo o escenario tecnológico supone todo un reto metodológico. En tercer lugar, el análisis de este trabajo se llevará a cabo a pesar de que la presencia de la IA en la actualidad al momento de escribir estas líneas no será comparable al del momento de su lectura, en cuanto al volumen de uso pero, sobre todo, al impacto de su utilización en las áreas que son objeto de estudio. Sin embargo, dada la potencialidad de su influencia, es urgente proceder con este tipo de análisis para identificar los problemas más importantes y urgentes, y aportar soluciones, o al menos contribuir con re-

* Trabajo elaborado en el marco del proyecto de investigación *Fortalecimiento de la democracia y el Estado de Derecho a través de la inteligencia artificial*, PID2021-122677NB-I00, financiado por el Ministerio de Ciencia e Innovación.

[1] En el marco del citado proyecto de investigación «Fortalecimiento de la democracia y el Estado de Derecho a través de la inteligencia artificial», se ha elaborado un glosario de términos esenciales para comprender la IA pero enfocada a los problemas del Estado de Derecho y la democracia. En la elaboración han participado constitucionalistas y expertos técnicos en IA con la pretensión de lograr un lenguaje común para sendas disciplinas y así facilitar el entendimiento entre ambas. Publicado en: https://www.idpucm.com/glosario/

flexiones sobre las mismas. Además, la velocidad del desarrollo tecnológico posiblemente no conceda (tanto) tiempo para debatir sosegadamente como sucedió con otros sucesos históricos. Nótese que aún hoy seguimos intentando llegar a acuerdos sobre cómo regular Internet, cuyo uso se popularizó hace un cuarto de siglo. Con la IA no tendremos tanto tiempo. La reflexión colectiva debe comenzar ya y debe hacerse siendo conocedores de la realidad, los retos y las herramientas constitucionales con las que contamos.

Hay debates ya abiertos de una densidad y complejidad que ameritan ser reflexionados desde muchos puntos de vista, pero que en última instancia nos exigen posicionarnos como comunidad política ante el hecho cada vez más constatable de que la tecnología no es totalmente neutral en el sentido de que, como señala COECKELBERGH, «la tecnología no es solamente un medio para alcanzar un fin, sino que también moldea ese fin»[2]. En consecuencia, la IA no es un asunto meramente tecnológico, pues «no es neutral en términos de política y poder. La IA es política de cabo a rabo» e impacta en cuestiones políticas clave. En definitiva, «La IA es política»[3]. Deben, por tanto, abordarse cuestiones esenciales como ¿qué papel va a jugar la IA en la sociedad? ¿Su utilización se va a imponer por sí sola como hasta ahora, como sucedió con Internet, o se va a decidir en y por la sociedad? ¿Qué criterios se usarán para responder a la pregunta anterior? Sin lugar a duda nos adentramos en un momento constituyente, pero como sociedad global, pues la IA es una herramienta horizontal y con unas características, como la autonomía y el aprendizaje, que implican una capacidad de transformación de nuestro mundo no vistas en las revoluciones tecnológicas anteriores.

A pesar de la mencionada revolución tecnológica, conocer adecuadamente los aspectos más relevantes y comunes del funcionamiento de las principales tecnologías que pueden encuadrarse bajo la etiqueta genérica de IA permite identificar las claves de los principales problemas suscitados en la actualidad en la sociedad algorítmica, como el abuso de poder fruto de un cada vez más desequilibrado reparto del mismo, una discusión pública cada vez más intoxicada que redunda en una menor calidad democrática y una decreciente observancia de los derechos fundamentales. Dichos problemas son de un calado fundacional.

Por eso, una disciplina imprescindible desde la que se debe abordar el impacto que está ocasionando una tecnología como la IA capaz de aportar un

---

[2] Coeckelbergh, Mark, *La filosofía de la inteligencia artificial. Una introducción*. Cátedra, 2ª edición, 2024. p. 16.

[3] *Ibid.*, pp. 17-18.

gran poder a quien la ostente y condicionar las relaciones de quienes componen la sociedad es el constitucionalismo.

Como sociedad decidimos ceder el ejercicio del poder en representantes, políticos y judiciales, que lo ejercen y toman decisiones en nuestro nombre. El Derecho Constitucional es el marco que regula las bases y los principios sobre los que se desarrolla esa decisión, organizando, limitando, impulsando y orientando su ejercicio. La irrupción de la IA, aunque sólo pueda vislumbrase, nos aboca a una nueva realidad social, una sociedad algorítmica, evolución de la sociedad digital donde hay nuevas realidades en el ejercicio del poder que quizá exijan un nuevo pacto social.

El centro del constitucionalismo y del Derecho Constitucional es el ser humano. Todos sus elementos como el pacto social, la toma de decisiones de manera democrática, el Estado de Derecho y los derechos humanos se conceptualizan, desarrollan e interpretan para que el ser humano sea autónomo y libre y, en consecuencia, pueda tomar sus decisiones libremente con plena capacidad de actuar y decidir (en la literatura anglosajona a esta capacidad y sus consecuencias se denomina *human agency*) en oposición a ejercicios tiránicos o abusivos del poder que anulen o socaven la naturaleza humana y su libre desarrollo.

Sin embargo, el desarrollo y aplicación de la IA tiene la potencialidad (y en muchos casos es ya realidad) de minar esta idea debido, no solo a su carácter autónomo y de autoaprendizaje, y a la ubicuidad de su presencia, sino a que esta tecnología opera con representaciones digitales de las personas, basadas en datos y modelos de decisión; representación que, sin embargo, no refleja la realidad de la persona.

En este trabajo se defiende que el constitucionalismo moderno y como consecuencia el Derecho Constitucional debe tener el claro cometido de evitar que el ser humano sea devaluado ante el uso creciente de una tecnología como la IA. En este sentido, la dignidad humana y la autonomía individual deben cobrar una radical centralidad en el carácter fundacional de nuestros sistemas sociales y jurídicos, por lo que demandan un reconocimiento y una protección más significada.

Al igual que una constitución es no solo un instrumento para limitar y controlar el poder, sino también una herramienta para el progreso social, la IA es una tecnología que ofrece potencialidades para transformar de manera positiva la sociedad, pudiendo ser utilizada para fortalecer instituciones, procesos democráticos y derechos.

Pero para atender a esos posibles usos positivos, esta tecnología puede exigir que categorías e instituciones clásicas del Derecho Constitucional sean reinter-

pretadas, modificadas o directamente ignoradas, exigiéndose a su vez la creación de nuevos conceptos.

En definitiva, en este trabajo se discute cómo la nueva realidad algorítmica en la que nos encontramos pone en tela de juicio el constitucionalismo clásico y se plantea un debate en torno a si deben abordarse, cuáles y con qué calado, cambios en el significado de instituciones y principios, llegando a cambios de paradigma en el constitucionalismo y en el Derecho Constitucional.

Estas reflexiones no pueden realizarse en abstracto. Es necesario realizar un recorrido por los diferentes escenarios clave de discusión y decisión colectiva y analizar críticamente el impacto que la IA está teniendo en la actualidad y el que presumiblemente va a tener. Por eso, en este trabajo se intercala el análisis de la realidad, normalmente al inicio de cada capítulo, con un estudio crítico sobre qué elementos deben mantenerse de manera irreductible y cuáles pueden ceder vía asimilación de nuevos contenidos y con qué intensidad, hasta el punto de tener que generarse nuevas categorías.

El objetivo de esta publicación no es agotar ninguno de los temas que se van a abordar, por la complejidad y extensión de los mismos, sino sobre todo describir el escenario que se abre en esta nueva realidad, identificar cuáles son los aspectos más problemáticos y más merecedores de atención por parte de un constitucionalista y sobre todo abordar aquellas cuestiones sobre las que pueden empezar a vislumbrarse alguna propuesta o solución o, al menos, tratar de acotarla. Pero hay temas que simplemente se dejan orientados por su excesiva complejidad y amplitud.

Para abordar este cometido, se lleva a cabo una lectura analítica y crítica desde el punto de vista de un constitucionalista de lo que puede entenderse como IA (Capítulo I), ofreciendo un concepto de IA y los elementos más esenciales para que un lego en la materia pueda comprender su funcionamiento y, en consecuencia, las amenazas y potencialidades que puede suscitar y ofrecer en el ámbito jurídico (Capítulo II), adoptando un necesario enfoque humanista de la IA (Capítulo III), analizando cómo la existencia y utilización de la IA está impactando en los principales elementos de los que se compone el constitucionalismo y el Derecho Constitucional, a saber, la necesidad un nuevo pacto social en la sociedad algorítmica (Capítulo IV), el impacto en la democracia (Capítulo VI), en el Estado de Derecho (Capítulo VII) y en los derechos humanos (Capítulo VIII).

El presente trabajo ha sido posible gracias a muchas personas, a quienes agradezco su ayuda y colaboración, del Departamento de Derecho Constitucional de la Universidad Complutense de Madrid, del Ministerio de Justicia y

del Consejo de Europa. Pero muy especialmente este trabajo ha visto la luz gracias a mi maestro, y sobre todo amigo, Rafael Bustos Gisbert, quien incansablemente me ha apoyado y animado a concluir estas páginas, contribuyendo de manera determinante a la calidad que en algún grado puedan contener. El sentimiento de gratitud y deuda es tan solo comparable con la ilusión que día a día pone en la forma de vivir una profesión que nos llena y nos hace feliz. No puedo concluir este párrafo de agradecimientos sin acordarme de mi familia, en especial de mi madre, y sobre todo de mi padre, a quien tanto echo de menos, e imagino disfrutando, leyendo estas páginas. El esfuerzo y sacrificio que ha supuesto este trabajo solo tiene sentido en compañía de Alejandro y de Carla, quienes no pueden hacerme más feliz. Y de Marta. No solo por su apoyo y cariño incondicional e inagotable, sino por su ilusión por formar el mejor equipo del mundo, a quien le dedico esta obra.

I

LA INTELIGENCIA ARTIFICIAL EN EL ESCENARIO CONSITUCIONAL

1. Nuevos elementos en la discusión constitucional

La historia atestigua claramente que el desarrollo y la aplicación de nuevas tecnologías siempre han producido cambios en la sociedad[1]. Si esos cambios fueron de suficiente entidad y calado, tuvieron reflejo en el ámbito político y, en consecuencia, en la organización del Estado.

Desde la aparición de Internet, la sociedad ha cambiado de manera radical: sus hábitos de consumo y de comunicación, la forma de relacionarse de sus miembros, el discurso público, el control político en las democracias... ¿Acaso no cambió la forma y relevancia del control parlamentario con la aparición de la entonces Twitter? Todo ha cambiado en apenas dos décadas. Como señala Rodatá, la «revolución de Internet» que diseña el mayor espacio público que la humanidad haya conocido jamás, se produce sin crear nuevas formas de relaciones institucionales marcando de esta manera las nuevas vías para un constitucionalismo global posible[2].

El cambio que propiciará la IA tendrá un calado mucho mayor y se producirá en mucho menos tiempo. De hecho, ya se está produciendo. Hay que ser conscientes, no obstante, que no son dos fenómenos separados, sino que el segundo cabalga a lomos del primero.

Por ello, es una obviedad afirmar que el desarrollo tecnológico tan acelerado que estamos experimentando en estas primeras décadas del siglo xxi en el marco de la cuarta revolución industrial[3] plantea cambios de paradigmas en la

---

[1] Brownsword, Roger, y Yeung, Karen, (eds), *Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes*, Hart Publishing, 2008.

[2] Rodatá, Stefano, *El derecho a tener derechos*, Trotta, 2012, p.22.

[3] Xu, M., David, J. M., & Kim, S. H., «The fourth industrial revolution: Opportunities and challenges», *International journal of financial research*, n. 9(2), 2018, pp. 90-95

noción del Estado nación y de las democracias liberales, tanto a nivel interno, en cuanto a la cohesión interna de la sociedad, como externo, con rápidas e importantes alteraciones geopolíticas[4].

La sociedad de la información ha evolucionado a una sociedad digital, algorítmica, en la que la masiva producción de datos y las diversas formas de tratamiento y explotación[5], como la acumulación de datos, minería de datos y análisis algorítmico tiene tremendas consecuencias a todos los niveles, desde la económica o empresarial (revolución industrial 4.0), hasta las instituciones y el Derecho[6].

La IA está mucho más presente en nuestras vidas de lo que somos conscientes y su uso continúa creciendo. Las posibilidades de desarrollo y evolución parecen no tener fin.

La tecnología no suele ser buena o mala por sí misma. Depende de su uso. Lo mismo sucede con la IA. Puede contribuir de manera muy significativa al progreso socioeconómico, pero su uso supone un riesgo en el disfrute de derechos o implica dejar de lado garantías jurídicas o institucionales. Pero también se ha de ser consciente de la realidad del desarrollo tecnológico y sus utilidades, pues es muy humano caer en la tecno-euforia o en el tecno-pesimismo. Tanto uno como otro deben evitarse, esforzándonos en realizar un ejercicio de realismo. La IA ha superado a los humanos en muchas actividades singulares, pero también de manera general en algunas cuestiones, como por ejemplo reducir o eliminar tanto sesgos como el factor aleatorio en la toma de una decisión. Como sintetiza Sunstein, «los algoritmos superan a los humanos, porque reducen o eliminan tanto el sesgo como el ruido (ruido es la toma de decisiones aleatoria), ya que pueden ayudar no solo a contrarrestar los sesgos humanos sino también a identificarlos.» Pero al mismo tiempo, «hay algunos problemas de predicción en los que los algoritmos no funcionarán bien; la razón radica en la ausencia de datos adecuados y, en cierto sentido, en lo que podríamos considerar la imprevisibilidad intrínseca de los asuntos humanos.»[7]

---

[4] Loesekrug-Pietri, André, «How Government and Civil Society Can Invent the State of the Twenty-First Century – The Case of Innovation», en FELDNER, Denise (ed.), *Redesigning Organizations. Concepts for the Connected Society*, Springer 2020, p. ix.

[5] Mayer-Schonberger, Viktor, y Cukier, *Big data. La revolución de los datos masivos*, Noema, 2013.

[6] Cohen, Julie., *Between Truth and Power. The Legal Construction of Information Capitalism*, Oxford University Press, 2020.

[7] Sunstein, Cass, «The use of algorithms in society», *The Review of Austrian Economics*, 4 de mayo de 2023, https://doi.org/10.1007/s11138-023-00625-z.

Tanto los aspectos positivos de la IA, como mejorar la transparencia en el uso del poder público y la participación política, como su uso para fines opuestos, como perturbar la democracia a través de interferencias en los procesos electorales, manipular la opinión pública, la economía, incluso influir definitivamente en la geoestrategia del orden mundial, tienen efectos en las vidas de las personas, e inciden sin duda en sus derechos en diferentes ámbitos. Sin embargo, no existe aún una plena consciencia de esta situación[8] y los problemas que genera, ni tampoco existe una supervisión adecuada.

Diariamente, todos interactuamos con sistemas de IA, que nos asisten en nuestra vida diaria, como sucede, por ejemplo, con las recomendaciones de Amazon o Netflix. Pero la IA no solo tiene la capacidad de recomendarnos con base en nuestras preferencias, sino que puede actuar y ejecutar decisiones de manera autónoma. Por ejemplo, el 75% de todas las transacciones de compraventa de acciones en las bolsas de EEUU se ejecutan de manera automática por algoritmos. El *flash crash* de la Bolsa estadounidense de 2010 fue motivado por un algoritmo malo[9]. Otro ejemplo, aún más alarmante, lo constituyen las armas autónomas con fuerza letal (*Lethal Autonomous Weapon System*, LAWS o *Autonomus Weapon System*, AWS). Esto que puede sonar a ciencia ficción, es ya una realidad[10].

Esta realidad suscita cuestionamientos constitucionales de carácter básico.

Sobran los ejemplos de retos constitucionales a nivel micro, como la elabo-

---

[8] Aunque se crea que la toma automatizada de decisiones por algoritmos es típico de países vanguardistas en lo tecnológico como Estados Unidos, o no muy respetuosos con los derechos fundamentales de sus ciudadanos como China, hay estudios, por ejemplo el realizado por la ONG Algorithmwatch que llaman la atención sobre el hecho de que en países de la UE ya se están utilizando, y van en aumento, algoritmos para la toma de decisiones automatizada o de apoyo a la toma de decisiones (ADM): el perfil de los solicitantes de empleo basado en sus correos electrónicos personales en Finlandia, la asignación de tratamiento para pacientes en el sistema de salud pública en Italia, la clasificación de los desempleados en Polonia, la detección de fraudes en la asistencia social en los Países Bajos, los sistemas de puntuación de créditos bancarios en muchos países de la UE son solo unos ejemplos, pero la gama de aplicaciones se ha ampliado a casi todos los aspectos de la vida diaria. Algorithmwatch: «Automating Society – Taking Stock of Automated Decision-Making in the EU»: https://algorithmwatch.org/en/automating-society/

[9] Benjamins, R., Salazar. I., *El mito del algoritmo. Cuentos y cuentas de la inteligencia artificial*, Anaya, Madrid, 2020, pp. 32 y 33.

[10] Véase el cerebro de defensa autónomo del ejército estadounidense Skyborg https://afresearchlab.com/technology/vanguards/successstories/skyborg ; https://www.larazon.es/tecnologia/20210530/rtibdefjtnflbft4yzbdz6xvlm.html

ración de perfiles personales o simplemente las decisiones predictivas, por mencionar dos problemas muy concretos e inmediatos. Pero desde el punto de vista macro, es indudable el cambio de paradigma que supone que el poder público ya no sea la fuente principal de amenaza al disfrute de los derechos fundamentales o de la democracia, sugiriendo reconsiderar o repensar categorías constitucionales tradicionales.

En primer lugar, se ha de reflexionar sobre la cada vez más generalizada toma de decisiones por algoritmos. Existe, pero irá creciendo de manera paulatina, una esfera de poder ejercida por sistemas automatizados, lo que se denomina como «*algocracia*»[11].

Dejando de lado los aspectos positivos, la reflexión constitucional ha de comenzar centrándose en los peligros de este poder, refiriéndose los más evidentes e inmediatos a cómo el rápido y huérfano de control desarrollo de sistemas de IA en campos relacionados con la administración pública, la justicia, o la política pueden producir decisiones erróneas o sesgadas, propiciando desigualdad o discriminación, decisiones injustas y socavando derechos fundamentales como libertad de expresión, intimidad o igualdad.

Pero el problema del poder ejercido por los algoritmos no se limita a los derechos fundamentales. La cuestión de fondo a la que hay que prestar más atención se refiere a la *legitimidad* de los sistemas automatizados que toman decisiones. El cuestionamiento de la legitimidad no solo trae causa de la opacidad de su funcionamiento, lo difícil o imposible que resulta explicar cómo ha alcanzado la decisión, sino que tampoco se sabe a veces por qué están tomando una decisión, es decir, por qué o en base a qué participa en el proceso de toma de decisión, o qué grado de participación han tenido en la misma. Es, sin duda, un problema de *legitimación democrática* al que quizá no se le está concediendo la suficiente importancia.

En segundo lugar, se ha de incidir en la no poco compleja problemática de las relaciones entre el Derecho Constitucional y el Derecho Privado. Sin duda, ha de prestarse mucha atención al papel tan relevante del *sector privado* que actúa como un reflejo de los poderes públicos.

El principal reto de las democracias constitucionales en la sociedad algorítmica podría consistir en limitar el auge de los poderes privados que intentan mostrar que su papel y sus decisiones tienen un sustrato democrático, pero en el

[11] Danaher, John, «The Threat of Algocracy: Reality, Resistance and Accommodation», *Philosophy & Technology*, n. 29, 2016, pp. 245-268.

fondo responden única y exclusivamente a determinaciones privadas[12]. Esto no implica intervenir en el mercado, sino definir un marco constitucional en el que los poderes públicos y privados estén sujetos a salvaguardias y procedimientos[13].

En tercer lugar, dada la exponencial evolución e innovación tecnológica en materia de IA, es difícil contar con marcos jurídicos, normativos o éticos de referencia actualizados. Y esta realidad, suscita una inquietante inseguridad jurídica, una indefinición de los límites y contenido de los derechos fundamentales, en donde cada vez más participan los poderes privados e incluso las máquinas, cuestionando la esencia del Estado de Derecho.»[14]

En efecto, hasta hace poco no había regulaciones jurídicas en vigor de carácter omnicomprensivas de este tipo de tecnologías, primando los compromisos éticos y los códigos privados de autorregulación[15]. Por norma general, los compromisos éticos se limitan a suscribir una autodeclaración, que describe una estructura interna con un comité de ética y un consejo asesor, pero su función y los poderes siguen sin estar claros, a pesar de que algunos de ellos están patrocinados por organismos políticos tan relevantes como la OCDE (Recomendación del Consejo sobre IA, adoptada el 22 de mayo de 2019[16]) o el G-7 y su *Hiroshima AI Process*, aprobado el 30 de octubre de 2023, que incluye 12 principios rectores sobre IA generativa y un Código de Conducta[17].

---

[12] Puede citarse como ejemplo paradigmático el intento de Meta de hacer pasar su *Oversight Board* o «Consejo Asesor Consultivo» (https://www.oversightboard.com/) una unidad interna y totalmente privada centrada en la moderación de contenido basada en las normas de política de uso de sus redes sociales, por una especie de verdadero tribunal que vela por la protección de los derechos humanos de los usuarios. Sánchez Frías, Alejandro, «El «tribunal supremo» de Facebook: ¿un nuevo paso hacia la justicia sin estado?», *Cuadernos de Derecho Transnacional*, vol. 12, n. 2, 2020, pp. 1386-1405.

[13] Pollicino, O., de Gregorio, G., «Constitutional Law in the Algorithmic Society», en Micklitz, H-W., Pollicino (et.al) (eds), *Constitutional Challenges in the Algorithmic Society*, Cambridge University Press, 2022, pp. 3-24, p. 24.

[14] Pollicino, O., de Gregorio, G., *op. cit.*, p. 9.

[15] Véase el imprescindible trabajo de JOBIN, A., Ienca, M., Vayena, E., «The global landscape of AI ethics guidelines», *Nature Machine Intelligence*, 1, 2019, pp. 389-399; Algortimwatch también está llevando a cabo un inventario de Guías éticas para la utilización de inteligencia artificial. Hasta el momento han inventariado 106, pero siguen trabajando en ello. Listado accesible en https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/

[16] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[17] https://www.soumu.go.jp/hiroshimaaiprocess/en/documents.html

Como ha afirmado la Asamblea Parlamentaria del Consejo de Europa, los instrumentos de *soft-law* y de autorregulación se ha comprobado que no son suficiente para hacer frente a los retos que la IA plantea y en particular los relacionados con los derechos humanos, la democracia y el Estado de Derecho[18]. Por eso, es cada vez más acuciante regular los sistemas algorítmicos y de IA, durante todo su ciclo de vida. Organizaciones internacionales, como el Consejo de Europa, han señalado insistentemente la necesidad de contar con un marco regulatorio para la inteligencia artificial, basado en la protección de los derechos humanos, el Estado de Derecho y la democracia[19].

Europa, precisamente, ha tomado la iniciativa regulatoria aprobando los dos primeros instrumentos de naturaleza jurídica vinculante[20], como es el Reglamento sobre IA aprobado por la Unión Europea[21] y la Convención Marco sobre IA y derechos humanos, democracia y Estado de Derecho, aprobada por el Consejo de Europa[22].

En definitiva, nos encontramos en el tránsito entre una sociedad de la información y una sociedad basada en datos y decisiones automatizadas. Por lo que «un análisis coherente de las adaptaciones institucionales que se derivan de la digitalización necesita abordar la dimensión técnica y sus implicaciones»[23]. Por eso, como señaló Hildebrant, «el reto al que se enfrenta el Derecho moderno es reinventarse a sí mismo en un entorno de informática preventiva sin renunciar a los logros fundamentales del Estado de Derecho.»[24] Estudiar cómo

---

[18] COMMITTEE ON POLITICAL AFFAIRS AND DEMOCRACY, PARLIAMENTARY ASSEMBLY, COUNCIL OF EUROPE, «Need for democratic governance of artificial intelligence», Doc. 15150, 24 de septiembre de 2020, de 2002, p. 6.

[19] Por ejemplo, «»Establishing a «legally binding instrument» for democratic governance of AI», Comisión Permanente de Asamblea Parlamentaria del Consejo de Europa, 22 de octubre de 2020, https://pace.coe.int/en/news/8059/establishing-a-legally-binding-instrument-for-democratic-governance-of-ai

[20] Dejando a salvo algunas normas nacionales reseñables a las que se hará referencia posteriormente como la Orden Ejecutiva del Presidente Biden (Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence), aprobada el 30 de octubre de 2023; o las normas aprobadas en China, a las que , por razones obvias, no se les prestará mucha atención.

[21] Publicado en el DOUE el 12 de julio de 2024.

[22] Abierto a la firma y ratificación el 5 de septiembre de 2024.

[23] Huether, Michael, «Old Institutions for a New World», en Feldner, Denise (ed.), *Redesigning Organizations. Concepts for the Connected Society*, Springer 2020, p. v.

[24] Hildebrandt, Mireille, *Smart Technologies and the End(s) of Law*, Edward Elgar, 2016, p. 17.

se está produciendo ese cambio es precisamente el objeto de este trabajo, pero acotado al Derecho Constitucional.

2. El papel del constitucionalista en este nuevo escenario algorítmico

¿Qué puede aportar aquí un constitucionalista?

Desde mi punto de vista, puede jugar dos papeles principales. El primero, identificar aquellas zonas constitucionales donde la IA puede tener un impacto significativo; el segundo, ayudar a desarrollar normativas y categorías de dimensión constitucional útiles para regular dichos impactos.

En primer lugar, desde un punto de vista constitucional, las implicaciones en los principales ámbitos constitucionales son evidentes.

Al ser una tecnología tan compleja, la primera exigencia para su uso es que los actores y organizaciones que la utilicen deben ser capaces de explicar cómo se llega a tomar una decisión. Sin embargo, los sistemas en cuestión pueden ser realmente ininteligibles. Tanto los que utilizan los sistemas de IA, como los responsables de regular su uso reconocen que no siempre los entienden del todo (produciéndose el fenómeno que se conoce como *blackbox* o caja negra[25]). E incluso los usuarios pueden no saber que están interactuando con esta tecnología. Esto entra de lleno en la esencia de las exigencias vertebradoras del Estado de Derecho, pues una de las principales formas de fomentar la protección de los derechos es garantizar que las personas puedan recurrir cuando algo les afecta negativamente.

Para ello, tienen que saber que se está utilizando la IA. Esto es más relevante aún cuando falta conciencia de las posibles implicaciones en materia de derechos que supone el uso de la IA. Existe una preocupación más extendida sobre la protección de datos, y también la no discriminación. Pero existe menos conocimiento respecto del peligro sobre otros derechos, como la dignidad humana, el acceso a la justicia y la protección del consumidor, entre otros[26].

---

[25] Hace referencia a la opacidad con la que algunos algoritmos de IA toman sus decisiones en función de la información introducida. Es común en tecnologías como redes neuronales, o *Deep learning*. La contraposición es la Inteligencia Artificial Explicable, que permite explicar la razón de las decisiones de un sistema.

[26] FUNDAMENTAL RIGHTS AGENCY *Getting the future right. Artificial Intelligence and Fundamental Rights*, 2020, p. 1.

Otra cuestión disputada es qué efecto está teniendo en las decisiones políticas, de políticos y de ciudadanía, la presencia de sistemas de IA en el escenario de los medios de comunicación y, más en general, en el funcionamiento del juego democrático. Es una realidad incontestable, que abordaremos más adelante, que la IA está jugando ya en la actualidad un papel político. Ha de prestarse atención también al efecto sobre la sociedad de la información hiperconectada en la que vivimos derivado de la concentración de un inmenso poder digital en pocas compañías privadas o en algunos Estados. Es imprescindible, por tanto, introducir en la IA el debate de la democracia y el debilitamiento del pacto social en el que se basan las sociedades occidentales.

Ya 2017, el Comité Económico y Social Europeo, en lo que se considera ampliamente el «informe inicial» de impacto social, identificó los ámbitos afectados más importantes, entre los que se encuentran: la seguridad; la ética; las leyes y la regulación; la democracia; la transparencia; la privacidad; el trabajo; la educación y la (des)igualdad[27]. Se ilustra claramente que esta tecnología tiene un impacto real en los derechos humanos, en la democracia y en el Estado de Derecho, los pilares esenciales sobre los que se edifican las sociedades europeas y occidentales.

En segundo lugar, por todo lo explicado anteriormente, es imprescindible que el jurista constitucionalista aborde esta dimensión de la realidad, trabajando de la mano con los desarrolladores que revisan el impacto potencial de los sistemas de IA, que tienden a centrarse en los aspectos técnicos, no jurídicos, y en la elaboración de las normas reguladoras del impacto de estos sistemas, aportando un contenido constitucional.

En este sentido, la visión constitucional puede aportar muchos aspectos positivos al debate de la regulación de los efectos de la IA.

En primer lugar, por ejemplo, el principio de responsabilidad exige distinguir entre regular la IA y regular el uso de esa tecnología. En el primer caso, los desarrolladores pueden escudarse en el efecto *Blackbox* y, por tanto, evitan la cuestión de la responsabilidad, para cuya experimentación se prevén ámbitos de *sandboxes*[28]. En el segundo caso, esa excusa no puede utilizarse,

[27] Comité Económico y Social Europeo, «Inteligencia artificial: las consecuencias de la inteligencia artificial para el mercado único (digital), la producción, el consumo, el empleo y la sociedad», INT/806, 2017, accesible en https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence-consequences-artificial-intelligence-digital-single-market-production-consumption-employment-and

[28] Entorno de prueba controlado que permite desarrollar y experimentar una tecnología controlando sus potenciales efectos perjudiciales.

pues su uso es algo que puede conocerse y se elige voluntaria y conscientemente.

En segundo lugar, la confianza está en el corazón de todas las reticencias hacia el uso de la IA. Como en política, la confianza, el *trust*, no es algo que se gane de manera inmediata, sino que es un proceso paulatino. Eso conecta con el importante tema de las certificaciones de sistemas de IA en los diferentes ámbitos en los que la ciudadanía no puede desconfiar, como, por ejemplo, en la administración de justicia[29].

En tercer lugar, y relacionado con los dos anteriores, el Derecho Constitucional exige en muchos sistemas de IA, o en etapas concretas de su desarrollo e implementación un enfoque «*human in the loop*» (HITL)[30], es decir, que una persona humana esté revisando siempre la decisión sugerida por el sistema. El ejemplo paradigmático sería el de las armas autónomas con fuerza letal.

En cuarto lugar, dado el potencial impacto que los sistemas inteligentes pueden tener en la vida de las personas, es importante diseñar esos sistemas con ese potencial impacto en mente. Por eso cada vez cobra más protagonismo la «IA centrada en el ser humano» (*human-centered AI*), una perspectiva sobre IA y *Machine Learning* que postula que los sistemas inteligentes han de ser diseñados teniendo en cuenta que forman parte de un sistema más amplio compuesto por actores humanos como usuarios, clientes, etc. De hecho, ya se empieza a utilizar la expresión «*human centered IA*» para referirse a los sistemas inteligentes que se diseñan teniendo en cuenta la responsabilidad social, por ejemplo, abordando cuestiones de equidad, responsabilidad y transparencia[31].

En quinto lugar, el Derecho Constitucional impone un marco conceptual, no solo de valores, sino también de técnicas y procesos que tratan de proteger al ciudadano de abusos de poder, se ejerza este en una época analógica o en la de la inteligencia artificial. No solo nos referimos en concreto a la tutela de los

---

[29] CEPEJ, Consejo de Europa, «Possible introduction of a mechanism for certifying artificial intelligence tools and services in the sphere of justice and the judiciary»: https://rm.coe.int/feasability-study-en-cepej-2020-15/1680a0adf4

[30] En la toma de decisiones autónomas, la intervención humana puede realizarse de distintos modos. *Human in the loop* (HITL) es en la que el ser humano está más presente; a continuación, en «human on the loop» (HOTL) el sistema toma decisiones, pero un humano supervisa por si hay errores; «human out of the loop» (HOOTL) quiere decir que el sistema toma decisiones autónomas sin ninguna intervención humana. Benjamins, R., Salazar. I., *op. cit.*, p. 306.

[31] Riedl, M., «Human-centered artificial intelligence and machine learning», *Human Behaviour and Emerging Technologies*, n. 1, 2019, pp. 33-36.

derechos fundamentales, sino a las garantías y principios conformadores del Estado de Derecho y de la democracia liberal. Como señaló BROWNSWORD, «es el ideal de legalidad junto con el Estado de Derecho lo que se interpone entre nosotros y un futuro tecnogestionado que nos priva de poder. La cuestión es, sin embargo, cómo debemos articular y aplicar estos ideales protectores.»[32] Por eso, el respeto constitucional desde el diseño de la IA es un tema al que nos referiremos de manera recurrente a lo largo de este trabajo.

En definitiva, los riesgos que estos avances puedan suponer para el ciudadano individual y para la sociedad en su conjunto son también una realidad. Se requiere, por tanto, que el Derecho regule este desarrollo, evitando perjuicios individuales y colectivos. De la protección de los derechos de los individuos y de su organización colectiva limitando el poder que puede ser potencialmente dañino se ocupa el Derecho Constitucional. De eso tratan estas páginas, en concreto centrándose en un mundo en el que la IA cobra cada vez más protagonismo.

Pero no puede proseguirse sin dejar claro que todo enfoque constitucional, como señala BALAGUER, está condicionado por una serie de rupturas constitucionales debido a la realidad digital[33], en el que el objeto de la constitución y del constitucionalismo está cambiando[34], por lo que «habrá que adoptar la propia constitución a las nuevas condiciones de un mundo nuevo que no se puede gobernar ya plenamente en los términos de la constitución analógica».[35]

---

[32] Brownsword, Roger, «Technological management and the Rule of Law», *Law, Innovation and Technology*, vol. 8 (1), pp. 100-140, 2016, p. 138

[33] Balaguer Callejón, Francisco, *La Constitución del algoritmo*, Fundación Manuel Giménez Abad, Zaragoza, 2022, pp. 21-24; 51-112.

[34] Sánchez Barrilao, Juan Francisco, «El derecho constitucional ante la era de ultrón: la informática y la inteligencia artificial como objeto constitucional», *Estudios de Deusto*, n. 64(2), 2016, pp. 225-258.

[35] Balaguer Callejón, Francisco, *op. cit.*, nota 38, p. 30.

II

UNA INTELIGENCIA ARTIFICIAL PARA IUSPUBLICISTAS. BASES CONCEPTUALES

1. Planteamiento. El problema de la definición

En la actualidad sigue sin haber una definición técnica totalmente consensuada del fenómeno tecnológico al que comúnmente nos referimos como inteligencia artificial sobre la que los expertos se pongan de acuerdo[1]. La dificultad para proponer una definición exacta estriba, no solamente en que se trata de un campo tecnológico en constante cambio y evolución, y, en consecuencia, el propio objeto de definición resulta difícil de aprehender, sino que bajo el término coloquial IA se conciben diferentes tecnologías y no todas pueden identificarse totalmente como IA; además, el concepto mismo de inteligencia es difícil de concretar, pues se compone y hace referencia a muchas facetas del comportamiento humano.

No obstante, y a efectos jurídicos en aras de los principios de legalidad y seguridad jurídica, varias instituciones han ofrecido su propia definición, pero con carácter finalista, es decir, no se ha definido la realidad tecnológica en sí, sino que se tratan de definiciones acotadas al ámbito de una norma jurídica y, por tanto, más limitadas en cuanto al objeto.

Baste, en esta parte inicial del trabajo, ofrecer una definición, o más bien una descripción muy simplificada, en la que poder pensar y basar el desarrollo del presente capítulo: «un sistema de IA es un software, un programa informático, que procesa datos y genera decisiones en base a los datos analizados»[2]. Esto es, desde la función que puede realizar, «los sistemas de IA son modelos algorít-

[1] Como ya señaló el Comisionado para Derechos Humanos del Consejo de Europa en su imprescindible Recomendación «Unboxing Artificial Intelligence: 10 steps to protect Human Rights» (Mayo 2019), no hay una definición consensuada de IA.

[2] Benjamins, R., Salazar. I., *op. cit.*, p. 38.

micos que desempeñan funciones cognitivas o perceptivas en el mundo que antes estaban reservadas a los seres humanos que piensan, juzgan y razonan.»[3]

En la misma línea se expresa la Asamblea Parlamentaria del Consejo de Europa, refiriéndose a «técnicas de procesamiento automatizado de datos que incrementan significativamente la habilidad de las máquinas para realizar tareas que requieran inteligencia, algo que ha sido casi exclusivo de los humanos hasta ahora. Equipados con algoritmos, máquinas modernas y robots pueden efectivamente aprender cosas nuevas, cambiar la forma en la que realizan tareas e implementar sus propias decisiones sin la intervención humana.»[4]

La Comisión Europea en su hoja informativa sobre Mercado Único Digital, se refirió a la IA como «sistemas que muestran comportamiento inteligente; al analizar su entorno pueden realizar una variedad de tareas con cierto grado de autonomía para alcanzar objetivos específicos»[5].

Más recientemente la UE describió la IA simplemente como «un conjunto de tecnologías que combinan datos, algoritmos y poder computacional»[6].

La Organización para el Desarrollo y Cooperación Económica (OCDE) y la UNCTAD (*UN Conference on Trade and Development*) han descrito la IA como «la habilidad de máquinas y sistemas de adquirir y aplicar conocimiento y llevar a cabo un comportamiento inteligente»

Todas estas aproximaciones o descripciones muestran que la IA comprende una variedad de tecnologías capaces de manipular objetos y tareas cognitivas, tales como sentir, razonar, aprender, y tomar decisiones. Pero no hay una única definición o incluso concepto, y será difícil que lo haya en virtud de la variedad de capacidades y procesos que alberga.

Encontramos, pues, aquí una primera dificultad. En aras de la seguridad jurídica, todo pensamiento y razonamiento jurídico requiere tener claro el ob-

---

[3] Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., Briggs, M. *Artificial intelligence, human rights, democracy, and the rule of law: a primer*, The Alan Turing Institute, 2021, p. 7.

[4] Committee on Social Affairs, Health and Sustainable Development, Parliamentary Assembly, Council of Europe, «Artificial intelligence and labour markets: friend or foe?», Doc. 15159, 5 de octubre de 2020, punto 4.

[5] European Commission, «Fact Sheet on a Digital Single Market. Artificial Intelligence for Europe», 24 de abril de 2018, https://digital-strategy.ec.europa.eu/en/library/factsheet-artificial-intelligence-europe, consultado el 18 de julio de 2025.

[6] Comisión Europea, Libro Blanco sobre la inteligencia artificial – un enfoque europeo orientado a la excelencia y la confianza, COM(2020) 65 final, 19 de febrero 2020. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065 consultado el 18 de julio de 2025.

jeto sobre el que se está trabajando. Esto no es tan sencillo con la IA como se acaba de exponer.

Quizás para intentar encontrar un concepto en una disciplina tan técnica y tan alejada de la jurídica, y en ausencia de una definición consensuada parece que el camino más adecuado sea comenzar entendiendo el propio objeto de estudio, es decir, ¿qué es la IA? ¿A qué nos referimos cuando hacemos referencia a la IA?

Ya sea al referirnos coloquialmente a ese término, o a cualquiera de las definiciones que se ofrecen por expertos, sean del mundo tecnológico, de la ingeniería, de las matemáticas, de la informática, o del mundo jurídico, se presume que el lector conoce decenas de años de evolución y desarrollo desde que el término inteligencia artificial fue acuñado por primera vez en la década de los 50 del siglo XX. Pero no suele ser lo habitual, con lo que se ignora el verdadero potencial de cambio de paradigma que supone esta tecnología.

Por tanto, para comprender adecuadamente el concepto de IA o cualquier definición es imprescindible familiarizarse con una serie de conocimientos esenciales para ayudar a perfilar y comprender los elementos que componen cualquier definición que se elabore o sugiera.

2. ¿Qué es la inteligencia artificial? Análisis de un concepto «coloquial» pero técnico para comprender su potencial transformador

Como se ha mencionado, hasta la fecha, no hay una definición de IA universalmente aceptada por la comunidad científica. El término, aunque se haya incorporado al lenguaje común, abarca un amplio y variado elenco de ciencias, teorías y técnicas, como el aprendizaje automático[7] (*machine learning*), el aprendizaje profundo[8] (*deep learning*), la ciencia de datos[9] (*data science*) e inclu-

[7] Subcampo de IA, que a su vez es un subcampo de Ciencias computacionales; algunas partes de *machine learning* puede pertenecer a estadística. Machine learning permite soluciones de IA que son adaptables. Una definición podría ser sistemas que mejoran su rendimiento en una tarea determinada con más y más experiencia o datos.

[8] Es una submateria de *machine learning* (que a su vez es una submateria de IA), que a su vez es una materia de ciencias computacionales. La profundidad («*depth*») del *Deep learning* se refiere a la complejidad de un modelo matemático;

[9] *Data science* (ciencia de datos): es un término que cubre diferentes subdisciplinas: incluye *machine learning* and *stattistics*, ciertos aspectos de ciencia computacional incluidos algoritmos,

so la robótica[10], cuyo objetivo es reproducir las capacidades de conocimiento de un ser humano.

Un primer intento de aproximación y de comprensión inicial a una realidad tan técnica y tan compleja podría llevarse a cabo a través de varios ejercicios, como por ejemplo, acudir a ejemplos de sistemas de IA que conocemos e incluso que utilizamos diariamente; descartar mitos y concepciones erróneas; aproximarse a su funcionamiento y elementos de los que se compone; identificar sus características principales definitorias; conocer los diferentes tipos y escuelas que la desarrollan; y, finalmente, comprender el carácter compuesto y cíclico de cualquier sistema de IA.

2.1 Ejemplos de uso diario de IA

Para los legos en la materia, una forma útil y sencilla de comenzar a conocer y conceptualizar la IA es acudir a aplicaciones reales con las que nos relacionamos diariamente y que utilizan IA, o simplemente aquellas que son muy conocidas y con las que estamos familiarizadas, ilustrando así diferentes aspectos de la IA[11].

En primer lugar, podrían mencionarse las herramientas de recomendación de contenidos que solemos utilizar a diario, a veces incluso sin ser conscientes de ello. Mucha de la información que nos encontramos a lo largo del día es información personalizada a través de IA. Por ejemplo, el contenido de nuestras redes sociales (*News Feed* de Facebook), anuncios publicitarios on line, recomendaciones musicales de Spotify, recomendaciones de Netflix o HBO… También personalizan contenidos los periódicos y compañías que tengan bus-

---

almacenamiento de datos y desarrollo de aplicaciones web. Es una disciplina práctica que requiere entender la materia en la que se aplica, por ejemplo, negocios o ciencia. Puede implicar un poco de IA. Podría definirse como «un campo que incluye elementos de varias disciplinas, como la informática las matemáticas, la estadística y las ciencias sociales, y generalmente se centra en la extracción de patrones de conjuntos de datos para responder o abordar una pregunta o un problema específico,» Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., Briggs, M. *op. cit.*, p. 7.

[10] Robótica: significa construir y programar robots para que puedan operar en escenarios complejos y reales. En cierto sentido, robótica es el último reto de IA puesto que requiere una combinación de todas las áreas de IA.

[11] Russel, S.J., Norving. P., Davis, E., *Artificial intelligence: A modern approach*, 3ª ed., Prentice Hall, 2010, pp. 28-29

cadores on line. Las portadas de periódicos pueden cambiar en función de las preferencias de cada usuario si están basados en IA. Esto tiene implicaciones muy relevantes en una sociedad de la información como en la que vivimos, al generarse filtros burbuja, factorías de *trolls*, *fake news*, y nuevas formas de propaganda, como se expondrá más detenidamente en el capítulo sexto sobre democracia.

En segundo lugar, sería útil apuntar aplicaciones variadas que utilizamos a diario y que funcionan con IA, como los navegadores Google Maps, Amazon, Gmail u otros programas que clasifican millones de correos como spam, Google Translate, y por supuesto los asistentes como Siri y Alexa.

En tercer lugar, podrían mencionarse los coches autónomos, pues requieren una combinación de técnicas de IA de muchos tipos: búsqueda y planificación para encontrar la mejor ruta entre dos puntos, visión computarizada para identificar obstáculos, y la toma de decisiones bajo incertidumbre para hacer frente a un entorno complejo y dinámico. Cada una de estas técnicas ha de trabajar con gran precisión para evitar accidentes. Estas mismas tecnologías se aplican a otros sistemas autónomos como los drones que vuelan, naves autónomas, robots de entrega, etc…

En cuarto y último lugar, podría apuntarse el tratamiento de imagen y vídeo. El reconocimiento facial ya es un producto utilizado en muchas aplicaciones de clientes, empresas y gobiernos, desde el mismo algoritmo que controla la toma de la fotografía hasta la organización de las fotos según las personas, el etiquetado automático en las redes sociales y el control de pasaportes. Técnicas similares pueden ser utilizadas para reconocer otros coches, matrículas y obstáculos alrededor de un coche autónomo, o para estimar las poblaciones de vida silvestre[12], sólo por nombrar algunos ejemplos. La IA puede utilizarse también para generar o alterar el contenido visual, como por ejemplo las aplicaciones para alterar fotos personales (tecnología *deep fake*), o las animaciones de Pixar donde los personajes animados replican los gestos hechos por actores humanos reales. Las implicaciones de esta tecnología pueden ser realmente importantes: cuando estas técnicas avancen y estén más ampliamente disponibles (véase la aplicación de Sora de Open AI), será fácil crear videos falsos de aspecto natural de eventos que son imposibles de distinguir de las imágenes reales, desafiando la noción de que «ver es creer». Al riesgo de esta tecnología nos referimos también en el capítulo sexto sobre democracia.

[12] https://valohai.com/success-stories/satellite-image-detection/ consultada el 18 de julio de 2025.

2.2 Mitos y concepciones erróneas

Descartar lo que no debe entenderse por IA es un ejercicio muy útil, pero sobre todo necesario, para combatir hasta eliminar el halo de misterio y casi misticismo al que se ha asociado a la IA. No nos referimos solo a la cuestionable elección de los términos que componen el binomio «inteligencia artificial» y a las problemáticas (y también ventajas) que suscita la elección de «metáforas» para su definición[13], sino a una serie de ideas que sencillamente no se corresponden con la realidad tecnológica.

La primera concepción errónea que se debe abordar es que no todo proceso automatizado implica la participación de un sistema de IA, sino que puede tratarse simplemente de una fórmula matemática compleja, o de un proceso robotizado sin ningún tipo de adaptación y «aprendizaje» del entorno que le rodea.

Un segundo mito que debe desterrarse es la antropomorfización de la IA, es decir, creer que detrás de un sistema de IA hay una esencia o «alma» humana, un sistema que actúa basándose en un razonamiento o ejercicio lógico como hubiera hecho un ser humano. No hay nada humano en la IA más allá de sus diseñadores y desarrolladores. El «pecado original» de este mito es la utilización de la palabra «inteligencia», que es una de las pocas características identificativas del ser humano. Si en lugar de «inteligencia artificial» se utilizara «sistema de computación a través de métodos estadísticos y matemáticos tremendamente avanzados», los prejuicios sobre esta tecnología estarían desprovistos de mucha mitología. De manera ilustrativa, y con pretensión de síntesis, los sistemas de IA se asemejan a cotorras estocásticas, es decir, sistemas que parece que toman decisiones como los humanos, pero que solo replican un comportamiento sin sentido ni lógica interna, de manera estocástica[14].

El tercer mito que se ha instalado en el ideario colectivo es que la IA no tiene intereses particulares y, en consecuencia, sería un gran gobernante para tomar decisiones en nuestro nombre, pues lo haría únicamente para satisfacer nuestros intereses. Sin embargo, todo sistema de IA se diseña por personas,

---

[13] Bustos Gisbert, Rafael, «El constitucionalista europeo ante la inteligencia artificial: reflexiones metodológicas de un recién llegado», *Revista Española de Derecho Constitucional*, n. 131, 2024, pp. 149-178, pp. 153-158.

[14] Bender, Emily M., (et.al.), «On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?», *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 610-623, https://doi.org/10.1145/3442188.3445922

con intereses particulares y para logar objetivos muy concretos que no tiene porqué identificarse con los de la generalidad de la sociedad, sino de pequeños grupos privilegiados con mucho poder y acceso a una tecnología exclusiva y cara.

El cuarto mito, muy relacionado con el anterior, es considerar que la IA es completamente objetiva porque su forma de proceder se basa en las matemáticas y la estadística, y estas disciplinas no pueden albergar ningún tipo de ideología, ni ser manipuladas de manera tendenciosa o partidista, puesto que «dos más dos siempre serán cuatro». Sin embargo, como ya expuso claramente O´-NEIL en un libro imprescindible[15], las matemáticas también pueden tener ideología puesto que se pueden utilizar para llegar a un resultado querido y no a otro. Además, los sistemas de IA suelen replicar sesgos que se esconden tanto en la formulación de los algoritmos como en la conformación de las bases de datos utilizadas para entrenar e implementar sistemas de IA.

El quinto y último mito que se abordará aquí hace referencia a la fe excesiva que se pone en este tipo de tecnología, considerando que puede llegar a solucionar todo tipo de problemas. Es lo que algunos autores han bautizado como «solucionismo tecnológico»[16]. Es indudable que la IA es una tecnología que ha conseguido mejorar la eficacia de muchos procesos, la detección de patrones ocultos y la toma de decisiones acertadas dado el estudio de ingentes cantidades de datos, de una manera cómoda y rápida. Muchos han querido extender este rotundo y fácil éxito a todos los problemas del ser humano. Pero no todos los problemas pueden ser solucionados por la IA, dependerá de la complejidad del problema, de la tecnología existente y de la voluntad de poner los medios disponibles a resolver ese problema; incluso la aplicación de la IA suscita problemas que contrarrestan esta afirmación y requiere una importante labor de diseño y desarrollo.

2.3. Funcionamiento y elementos de la IA

Un tercer ejercicio consistiría en describir, de manera muy simplificada, quizá en exceso, su funcionamiento, permitiendo al mismo tiempo identificar

[15] O`Neil, Cathy, *Armas de destrucción matemática. Cómo el big data aumenta la desigualdad y amenaza la democracia*, Capitán Swing, 2017.

[16] Morozov, Evgeny, *To Save Everything, Click Here: The Folly of Technological Solutionism*, Public Affairs, New York, 2013.

sus principales elementos[17]. Para ello es imprescindible partir de un concepto que utilizan los expertos en IA y que marca el funcionamiento de todo sistema de IA: la racionalidad. Un sistema de IA es, ante todo *racional*, más que inteligente, dada la vaguedad y mayor complejidad de este concepto, como se abordará más adelante. La racionalidad es «la capacidad de elegir la mejor acción posible para alcanzar un objetivo determinado, dados determinados criterios que es necesario optimizar y teniendo en cuenta los recursos disponibles.»[18] El siguiente paso es describir cómo alcanza un sistema de IA la racionalidad, es decir, cómo es capaz de elegir la mejor acción posible. A efectos pedagógicos, se puede estructurar el proceso en tres fases: una primera de recopilación de datos, una segunda de tratamiento de información, y una tercera de toma de decisiones. Es decir, y hablando en términos generales, un sistema de IA debe contar con tres capacidades fundamentales: percepción, razonamiento/toma de decisiones y acción, que pueden ser reconducidas a la capacidad de razonamiento y aprendizaje.

La primera fase consiste en percibir y analizar el entorno en el que se encuentra inmerso el sistema de IA a través de sensores o dispositivos de captación de información, que pueden tratarse de cámaras, micrófonos, un teclado, un sitio web, o sensores de cantidades físicas, como de temperatura, distancia, sensores táctiles, etc. Es una obviedad resaltar la importancia de dotar al sistema de IA de sensores adecuados para percibir los datos para lograr el objetivo establecido para el sistema de IA.

La segunda fase se centra en el razonamiento y tratamiento de la información. Los datos recabados por los sensores son procesados y deben convertirse en información que el sistema pueda comprender. Esa información debe ser procesada y proponer la mejor acción de acuerdo con el objetivo fijados. Por tanto, es crucial transformar los datos en conocimiento.

La tercera fase es llevar a cabo esa acción propuesta, pudiendo así modificar o influir en su entorno.

Respecto del aprendizaje, existen diferentes técnicas y variedades, que no serán explicadas puesto que su complejidad excede con creces el propósito de este trabajo[19].

---

[17] Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, *Una definición de la Inteligencia Artificial: Principales capacidades y disciplinas científicas*, abril 2019, pp. 2-4, accesible en https://agatadata.com/wp-content/uploads/2024/11/Una-definicion-de-la-inteligencia-artificial-2019-Comision-Europea.pdf, consultado el 18 de julio de 2025.

[18] *Op. cit.*, p. 1.

[19] Para una explicación nada técnica, véase, *op. cit.*, pp. 3-5

2.4. Características definitorias de la IA

Un cuarto ejercicio para aproximarnos a la realidad o concepto de IA es identificar las principales propiedades características de todo sistema de IA. Sin ánimo de exhaustividad, podrían señalarse dos imprescindibles: autonomía y adaptabilidad.

La *autonomía* es la capacidad o habilidad de realizar tareas en entornos complejos sin la guía constante de un usuario; o, en palabras de la OCDE, «el grado en que un sistema puede aprender o actuar sin intervención humana tras la delegación de autonomía y automatización de procesos por parte de los humanos»[20]. Por su parte, la Organización Internacional de Normalización (ISO) la define como la «característica de un sistema que es capaz de modificar su ámbito de uso u objetivo previsto sin intervención, control o supervisión externos»[21].

Por su parte, la *adaptabilidad* es, a grandes rasgos, la capacidad de mejorar el rendimiento aprendiendo de la experiencia. Hace referencia a la capacidad de seguir evolucionando tras su desarrollo inicial, modificando «su comportamiento a través de la interacción directa con la entrada y los datos antes o después de su despliegue»[22].

En síntesis, todo sistema tecnológico que contenga estas dos habilidades podría caracterizarse como «inteligente».

2.5. Tipos de IA

Un siguiente paso y que además ayuda a desterrar tópicos erróneos, sería distinguir entre los diferentes tipos de IA, principalmente entre IA específica, débil o estrecha (*narrow*, ANI por sus siglas en inglés), por un lado, e IA fuerte o general (o *general*, AGI) por otro.

---

[20] OECD (2024), «Explanatory memorandum on the updated OECD definition of an AI system», OECD Artificial Intelligence Papers, No. 8, OECD Publishing, Paris, https://doi.org/10.1787/623da898-en

[21] Apartado 3.1.5 de la Norma ISO/IEC 22989:2022(e), Information technology — Artificial intelligence — Artificial intelligence concepts and terminology.

[22] Algunos ejemplos son un sistema de reconocimiento de voz que se adapta a la voz de un individuo o un sistema de recomendación de música personalizado. OECD (2024), «Explanatory memorandum on the updated OECD definition of an AI system», OECD Artificial Intelligence Papers, No. 8, OECD Publishing, Paris, https://doi.org/10.1787/623da898-en

Por IA general se hace referencia a una máquina o sistema que está diseñada para realizar la mayoría de las actividades que pueden llevar a cabo los seres humanos[23], es decir, que puede abarcar cualquier tarea intelectual, una réplica de las capacidades y habilidades de un ser humano[24]. La IA general equivaldría a una «mente» que es genuinamente inteligente y consciente de sí misma. Obviamente, en la actualidad no existe semejante inteligencia más allá de las películas de ciencia ficción, y todos los expertos coinciden en señalar que tardará en lograrse algo semejante, si es que se llega a lograr[25].

Lo que sí existe en la actualidad, y es cada vez más generalizada, son métodos o programas que entran dentro del concepto de IA específica o débil. Los sistemas de IA específicos solamente pueden realizar una tarea concreta o un número limitado de ellas, y hoy día superan ampliamente la capacidad humana, como por ejemplo jugar al ajedrez. Todas las tecnologías usadas hoy entran dentro de la IA específicas; la IA fuerte, en cambio, entra dentro de la ciencia ficción. En los últimos 50 años la IA general no ha avanzado, al contrario que la IA específica, que ha avanzado muchísimo. La IA específica es lo que realmente tenemos, es decir, sistemas que exhiben comportamientos inteligentes a pesar de ser «meros» ordenadores.

No obstante, en los últimos años cada vez se producen más avances en sistemas multimodales, es decir, sistemas que combinan diferentes sistemas de IA débiles, avanzando por tanto hacia modelos con vocación cada vez más general. El avance vertiginoso en la IA de la década de los años 20 del siglo XXI puede convertir en inexactas o incluso falsas las afirmaciones precedentes sobre la IA general.

2.6. Escuelas de desarrollo de la IA

Otro aspecto interesante sería distinguir entre los enfoques y escuelas a partir de los cuales se han desarrollado diferentes formas de IA.

---

[23] Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, *op. cit.*, p. 5.

[24] Otra visión de esta distinción es considerar la IA débil la hipótesis de que las máquinas pueden actuar como si fueran inteligentes; en cambio la afirmación de que las máquinas que lo hacen están realmente pensando (no sólo simulando el pensamiento) se denomina hipótesis fuerte de la IA, Russel, S.J., Norving. P., Davis, E., *op. cit.*, pp. 1020-1033.

[25] Por todos, véase el imprescindible Larson, Erik J., *El mito de la inteligencia artificial. Por qué las máquinas no pueden pensar como nosotros lo hacemos*, Shackleton Books, 2022.

La IA es una disciplina con numerosas ramas de conocimiento como se ha apuntado ya, que se nutren de las dos grandes escuelas de pensamiento *top-down* y *bottom-up*[26]. Estas dos aproximaciones a la IA son muy distintas conceptualmente y han generado una especie de enfrentamiento entre ambas:

— la primera está basada en la *lógica*, adopta un enfoque simbólico-lógico (*top-down*, originariamente llamado *neat*) y genera sistemas basados en el conocimiento;
— la segunda está basada en datos y adopta un enfoque conexionista, (*bottom-up*, originariamente conocido como *scruffy*), que genera sistemas «inteligentes» de aprendizaje automático.

El primero, *top-down*, o de arriba abajo, que forman parte de la escuela de la «IA simbólica», postula que «las máquinas para razonar debían seguir un conjunto de reglas predefinidas y unos principios de la lógica. La idea es programar en la máquina el conocimiento que poseemos los humanos, de forma que después, aplicando las reglas, que también han sido enseñadas previamente, el ordenador pueda derivar un conocimiento nuevo». Incluye áreas como la teoría de juegos, la lógica, la representación del conocimiento, la teoría del aprendizaje, etc.[27].

Por tanto, estos sistemas tienen dos componentes esenciales: en primer lugar, una *base de conocimiento* y, en segundo lugar, un «*motor de inferencia*», que extrae un juicio o conclusión a partir de los hechos contenidos en dicha base[28].

La *base de conocimiento* contiene hechos y relaciones relevantes para el propósito final del algoritmo, y a su representación se le conoce también por *ontología*. Por ejemplo, en un sistema de IA que diagnostica automóviles, la base de conocimiento contendría información sobre los componentes de un coche, cómo están montados, el lugar exacto, la función que desempeña y su funcionamiento, sus fallos más frecuentes y causas, etc. En el caso de un sistema de IA médico, la base de conocimiento consistiría en las partes del

[26] Oliver, N., *Inteligencia Artificial, naturalmente. Un manual de convivencia entre humanos y máquinas para que la tecnología nos beneficie a todos*, Ministerio de Asuntos Económicos y Transformación Digital, 2020, pp. 37.

[27] *Ibid.*, p. 38.

[28] Benjamins, R., Salazar. I., *op. cit.*, p. 40.

cuerpo, los órganos relevantes, cómo funcionan y se relacionan entre sí, dónde se encuentran en el cuerpo, etc. Usualmente, estas bases de conocimiento son creadas por los llamados «ingenieros de conocimiento», que extraen la información relevante de la literatura o lo adquieren mediante entrevistas con expertos[29].

Además de esta base de conocimiento, que representa todos los hechos que el ingeniero cree que deben ser tenidos en cuenta, y las relaciones entre ellos, estos sistemas basados en conocimiento «razonan» a través del *motor de inferencia* con los hechos, para dar como resultado «nuevos hechos y conclusiones.» Un ejemplo clásico de razonamiento lógico serían los silogismos de diferentes tipos, compuesto por premisa mayor, premisa menor y conclusión. Existen estrategias diferentes para encadenar estas reglas, como «razonamiento hacia atrás», que comienza por la conclusión y aplica reglas hasta que solo quedan hechos, o el «razonamiento hacia adelante», que comienza por los hechos y ejecuta las reglas hasta haber llegado a una conclusión válida. Estos algoritmos se suelen llamar «sistemas basados en conocimiento», «sistemas expertos» o «sistemas basados en reglas».[30]

El segundo enfoque, *bottom-up*, de abajo a arriba, «proponía que la Inteligencia Artificial debía inspirarse en la biología, aprendiendo a partir de la observación y de la interacción con el mundo físico, esto es, de la experiencia. Según este enfoque, si aspiramos a crear IA debemos proporcionar a los ordenadores observaciones de las que aprender. Esto conlleva entrenar algoritmos a partir de miles de ejemplos de los que queremos que aprendan.» Esta escuela incluye la percepción computacional, que abarca el procesamiento de imágenes, vídeos, texto, audio y datos de otro tipo de sensores; el aprendizaje autónomo o *machine learning*, los métodos de búsqueda (de texto, imágenes, vídeos), los sistemas de agentes, la robótica, los sistemas de recomendación y personalización, las inteligencias social y emocional computacionales, entre otros[31].

La fórmula que más se está usando en la actualidad para su desarrollo es el llamado *machine learning* o aprendizaje automático[32]. En este caso, al contrario

---

[29] Benjamins, R., Salazar. I., *op. cit.*, pp. 40-41.

[30] *Ibid.*

[31] Oliver, N., *op. cit.*, pp. 38 y 39.

[32] «Un tipo de computación que se utiliza para encontrar patrones en los datos y hacer predicciones de un resultado para una instancia particular. El término «aprendizaje» es un poco engañoso, ya que el ordenador no aprende de la misma manera que los humanos. En cambio, el ordenador es capaz de encontrar similitudes y diferencias en los datos a través del ajuste repetiti-

del cerebro humano, la IA necesita grandes cantidades datos para realizar el análisis (el aprendizaje) que le lleva a sus conclusiones. La IA, por ahora, se alimenta de bases de datos que nosotros le proporcionamos, es decir, su aprendizaje se realiza a través de la «ingesta» masiva de datos. No en vano, uno de los principales obstáculos que impidió a esta tecnología poder desarrollarse desde sus inicios, al comienzo de los años 50, fue la escasez de datos. En la actualidad, la generalización en el uso de Internet ha solucionado este problema generando un Big Data, esto es, con un crecimiento exponencial de datos, un incremento en la potencia de los ordenadores y el aumento de la capacidad de almacenaje de los datos[33].

---

vo de sus parámetros (a menudo llamado «entrenamiento»). Cuando los datos de entrada cambian, los resultados también cambian en consecuencia, lo que significa que el ordenador aprende a detectar nuevos patrones. Esto se consigue aplicando una fórmula matemática a grandes cantidades de datos de entrada para producir un resultado correspondiente.», Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., Briggs, M., *op. cit.*, 2021, p. 7.

[33] El concepto big data abarca dos esferas: por un lado, la ingente cantidad de datos, de diferentes tipos, procedencias y con unas características determinadas; por otro lado, la utilización de programas algorítmicos o matemáticos capaces de abarcar semejante volumen de datos para establecer correlaciones o patrones entre esa cantidad de datos utilizados. De hecho, a pesar de que el nombre big data puede hacer más referencia a la vasta cantidad de datos a la que se refiere, hay autores que destacan mucho más las capacidades potentísimas que esta tecnología tiene para buscar, agregar, y hacer referencias cruzadas dentro de grupos gigantescos de datos (por ejemplo, Boyd, d., & Crawford, K., «Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon», *Information, Communication & Society*, vol. 15(5), 2012, pp. 662-679). Las principales características definitorias del big data son conocidas como las 5 Vs: ***volumen***, es decir, la ingente cantidad de datos a la que se hace referencia cuando se menciona el big data, pero también hace referencia a los mecanismos para poder almacenar, procesar y tratar cantidades tan vastas de datos; ***variedad***: la gran diversidad de fuentes y tipos de datos que se reúnen en una tecnología de big data. Los datos que son tratados mediante esta tecnología pueden ser estructurados (fechas, números…), semiestructurados y no estructurados (video, imagen, voz, sonido, emails, contenido de redes sociales, texto), siendo la mayor parte de la información actual no estructurada. La procedencia de estos datos también es muy variada, pueden ser generados por personas, transacciones de datos, navegación por la web, IoT, etc; puede hablarse también de otras categorizaciones como *hard data*, *soft data* y *metadata*; ***velocidad*** con la que los datos son generados, acumulados y procesados. La captación de los datos, su tratamiento y análisis pretende llegar a realizarse a tiempo real, minimizando el periodo de latencia. Tampoco nunca antes se había podido hacer de manera tan rápida. Esto significa también que la utilidad, la importancia y el valor de los datos evolucionan en tiempo real, en línea con la entrada de nuevos datos; ***veracidad*** o fiabilidad de los datos: esta característica hace referencia a cómo de confiables son las fuentes de los datos que se manejan, o los mismos datos; por último, ***valor***: la utilidad de cada unidad o elemento de dato depende de la

No obstante, para que sean útiles para los sistemas de IA, los datos han de cumplir una serie de requisitos, no siendo útiles aquellos que no sean veraces, los que pertenezcan a situaciones muy específicas, ya sea temporales o culturales, y a los sacados de contexto. Los datos, por sí solos, no generan conocimiento, pero sí su análisis y procesamiento. El peligro en la toma de decisiones por parte de los algoritmos es que, si un eslabón en esta cadena de proceso está alterado, acaba afectando a su totalidad[34].

El algoritmo o el sistema utiliza los datos para entrenarse, como lo haría cualquier persona, para conseguir su objetivo: tomar la mejor decisión posible. Cuanto más se entrene y, por tanto, de cuantos más datos disponga, mejor llegará a su objetivo, es decir, a una decisión. «El modelo compara, analiza y extrae patrones. Así puede clasificar, discriminar y agrupar en distintas categorías útiles. Gracias a estos patrones es capaz de «afrontar» situaciones nuevas. Las compara con las que ya «ha vivido» similares y extrapola. Así, también puede realizar «predicciones», a través de cálculos de probabilidad, con un alto índice de acierto, siempre en función de su nivel de entrenamiento y de la fiabilidad de los datos con los que ha sido entrenado.»[35]

Dentro de ese enfoque *bottom-up*, a mediados de los años 90 y especialmente en la última década, las técnicas de aprendizaje estadístico por ordenador (*statistical machine learning*) ha experimentado un avance muy significativo, y junto con la confluencia de otros tres factores como el acceso a cantidades ingentes de datos – big data -, la disponibilidad de procesadores muy potentes a bajo coste, y el desarrollo de redes neuronales profundas y complejas (*deep*

---

cantidad de otros datos con los que se pueda conectar o correlacionar, más que la densidad de información que tenga. Incluso los datos que tengan muy poca información (por ejemplo, datos anónimos) se convierten más útiles cuando más numerosos sean. Por tanto, el potencial valor (por supuesto, también económico) de los datos es latente, depende de poder conectar cada dato con otros muchos datos. Es lo que se denomina el efecto red (*network effect*), que podría ser definido como el efecto por el cual el valor real de algo, de una técnica o de un producto depende directamente del número de sus usuarios. Aplicado al big data, el efecto red produciría que el valor real de los datos depende de la cantidad de otros datos recopilados con lo que se pudieran agregar. Véase, Rouvroy, A., «Of Data and Men». Fundamental Rights and Freedoms in a World of Big Data», T-PD-BUR(2015)09REV, Council of Europe, Strasbourg, 11 January 2016, pp. 5-10; Kitchin, R., *The Data Revolution. Big Data, Open Data, Data Infrastructures & Their Consequences*, Sage, 2014; Kuner, C., Cate, F.H., Millard, C., «The challenge of big data for data protection», *International Data Privacy Law*, 2012, vol. 2(2), pp. 47-49; Mayer-Schönberger, V., Cukier, K., *op. cit*.

[34] Benjamins, R., Salazar. I., *op. cit.*, p. 38.

[35] Benjamins, R., Salazar. I., *op. cit.*, p. 39.

learning) han instalado hoy día a la IA «en una ·primavera perpetua», en palabras del profesor de la Universidad de Stanford Andrew Ng[36].

2.7. La IA se desarrolla en un ciclo de vida

El último, pero no por ello menos importante, de los ejercicios que se propone para entender qué es la IA aborda uno de los conceptos más importantes y que más información puede aportar: el ciclo de vida.

Como define la ISO, el ciclo de vida de un sistema de IA «describe la evolución de un sistema de IA desde su inicio hasta su retirada»[37]. Más en detalle puede definirse como el proceso compuesto por pasos o etapas que cualquier persona u organización ha de abordar si pretende que un sistema de IA solucione un problema, tarea o conjunto de tareas, desde el primer momento en el que se concibe hasta su retirada tras haber sido puesto en marcha en la vida real. Consiste, por tanto, en una progresión secuencial de tareas y decisiones que impulsan el desarrollo y la implantación de soluciones de IA.

Conocer y entender que el ciclo de vida de un sistema de IA se compone de diferentes etapas permite a su vez identificar y señalar qué actividades concretas pueden ser más proclives a la vulneración de derechos o de otros intereses jurídicamente protegidos, y por tanto requerir de la adopción de cautelas especiales.

El ciclo de vida de un sistema de IA puede dividirse en tres etapas principales, que a su vez puede contener diferentes pasos: la primera de diseño (*design*), la segunda de desarrollo (*development*) y la tercera de despliegue (*deployment*)[38].

La etapa de *diseño* comprende varios pasos, como la planificación del proyecto, la formulación del problema, y la extracción u obtención de datos y su análisis.

La planificación del proyecto exige, en primer lugar, fijar los objetivos del proyecto, además de permitir reflexionar sobre las cuestiones éticas, socioeco-

---

[36] Oliver, N., *op. cit.*, p. 47.

[37] Apartado 6.1 de la Norma ISO/IEC 22989:2022(e), Information technology — Artificial intelligence — Artificial intelligence concepts and terminology

[38] Véase, por todos, Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., Briggs, M. *op. cit.*, pp. 9-11. La ISO sigue un esquema similar, señalando diferentes etapas: concepción (inception), diseño y desarrollo, verificación y validación ejecución, funcionamiento y monitorización, validación continua, re-evaluación y retirada, apartados 6.2.2 a 6.2.9 de la Norma ISO/IEC 22989:2022(e), Information technology — Artificial intelligence — Artificial intelligence concepts and terminology.

nómicas, jurídicas y técnicas antes de invertir recursos en el desarrollo del sistema. A continuación, debe determinarse qué problema se va a abordar, qué datos de entrada se necesitan y con qué fin. Obviamente, se debe tener en cuenta las implicaciones éticas y jurídicas de los usos de los datos y proporcionar una descripción exhaustiva de las consecuencias previstas e imprevistas del uso. Esta etapa comprende los procesos mediante los cuales se *recopilan* datos para el problema en cuestión: puede realizarse mediante la extracción de datos, que puede implicar procesos de web *scraping,* de registro de datos mediante encuestas o metodologías similares; o puede realizarse también mediante la obtención a través de acuerdos legales para obtener conjuntos de datos ya existentes. Consecuentemente, en esta etapa se podría comenzar a inspeccionar los datos, abordando un análisis exploratorio exhaustivo de datos, que implica comprender la composición de los datos pudiendo surgir preguntas como si los datos son incompletos, atípicos o desequilibrados.

La etapa de *desarrollo* implica también llevar a cabo una serie de pasos, como el procesado previo o preprocesamiento de datos, la selección del modelo y su entrenamiento, su testeo y validación y, por último, su evaluación.

El procesado previo o preprocesamiento de los datos suele ser el paso que más tiempo consume en la fase de desarrollo del ciclo de vida de la IA; incluye tareas que contribuyen a la alimentación del proceso de entrenamiento del modelo, como la limpieza de datos y su transformación en un formato propicio para la modelización. También se debe seleccionar el modelo que sirva al problema determinado en la etapa de diseño, teniendo en cuenta factores como los tipos de datos, su cantidad y su disponibilidad, que serán utilizados para entrenar al modelo. Tras el entrenamiento, el modelo se ajusta y se prueba con datos «no vistos». A continuación, el modelo se prueba con datos desconocidos para imitar la aplicación en el mundo real y confirmar el rendimiento y la precisión. Por último, una vez que se entrene, valide y pruebe el modelo, se debe evaluar incluyendo medidas de rendimiento, evaluaciones de impacto e información sobre el flujo de trabajo.

En la etapa final de *despliegue*, se lleva a cabo la ejecución del modelo (*model implementation*), el entrenamiento por parte del usuario (*user training*) y el seguimiento o monitorización (*monitoring*).

La última etapa del ciclo de vida de un sistema de IA consiste en implantar o desplegar el modelo entrenado *en el mundo real*. Una implementación eficaz permite incorporar el modelo a un sistema más amplio. Los nuevos datos obtenidos en la vida real son procesados por el modelo implementado para servir al propósito determinado en la fase de diseño. Los responsables de la aplicación