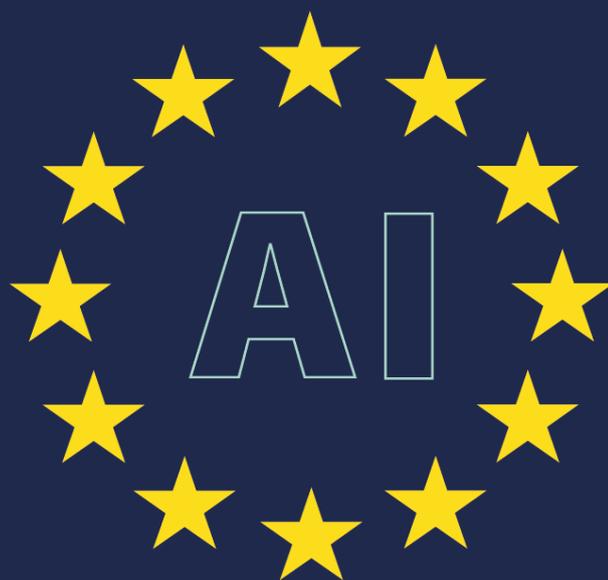


**GRUPO INDEPENDIENTE DE EXPERTOS DE
ALTO NIVEL SOBRE
INTELIGENCIA ARTIFICIAL**

CREADO POR LA COMISIÓN EUROPEA EN JUNIO DE 2018



**UNA DEFINICIÓN DE LA INTELIGENCIA
ARTIFICIAL:
PRINCIPALES CAPACIDADES Y
DISCIPLINAS CIENTÍFICAS**

Una definición de la inteligencia artificial: Principales capacidades y disciplinas científicas

Grupo de expertos de alto nivel sobre inteligencia artificial

Exención de responsabilidad y uso de este documento: la descripción y definición que siguen acerca de las capacidades de la IA y las áreas de investigación existentes en este campo constituyen una simplificación considerable del estado de la situación. La finalidad de este documento no es definir de manera precisa y exhaustiva todas las técnicas y capacidades de la IA, sino ofrecer una breve descripción de la visión común que aplica el Grupo de expertos de alto nivel en esta disciplina al elaborar sus entregables. Los autores confían, no obstante, en que este documento pueda servir también como un punto de partida útil para el aprendizaje por parte de las personas que no sean expertas en IA, de modo que, a partir de la información que se expone en él, puedan llevar a cabo una reflexión más amplia y profunda sobre la IA y conocer mejor esta disciplina y la tecnología conexas.

El Grupo de expertos de alto nivel sobre IA es un grupo de expertos independientes constituido por la Comisión Europea en junio de 2018.

Contacto Nathalie Smuha - Coordinadora del Grupo de expertos de alto nivel sobre IA
Correo electrónico CNECT-HLG-AI@ec.europa.eu

Comisión Europea
B - 1049 BRUSELAS

Documento publicado el X de abril de 2019.

Se publicó un primer borrador de este documento el 18 de diciembre de 2018, junto con el primer borrador de las directrices éticas del Grupo de expertos de alto nivel sobre inteligencia artificial para una IA fiable. El documento se revisó teniendo en cuenta los comentarios recibidos a través de la Alianza europea de la IA y la consulta abierta sobre el proyecto de directrices. Los autores desean expresar su sincero agradecimiento a todas las personas que realizaron aportaciones al primer borrador del documento.

Ni la Comisión Europea ni cualquier persona que actúe en su nombre serán responsables del uso que pudiera hacerse de esta información. El contenido de este documento de trabajo es responsabilidad exclusiva del Grupo de expertos de alto nivel sobre inteligencia artificial. Si bien el personal de la Comisión colaboró en la elaboración de este documento, los puestos de vista expresados en él reflejan la opinión del Grupo de expertos de alto nivel sobre inteligencia artificial; en ningún caso se considerará que representan una posición oficial de la Comisión Europea.

Encontrará más información sobre el Grupo de expertos de alto nivel sobre inteligencia artificial en línea (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>).

La política de reutilización de los documentos de la Comisión Europea está regulada por la Decisión 2011/833/UE (DO L 330 de 14.12.2011, p. 39). Para cualquier uso o reproducción de fotografías u otro material no sujeto a los derechos de autor de la UE, debe solicitarse permiso directamente a los titulares de los derechos de autor.

UNA DEFINICIÓN DE LA INTELIGENCIA ARTIFICIAL: PRINCIPALES CAPACIDADES Y DISCIPLINAS CIENTÍFICAS

Partimos de la siguiente definición de inteligencia artificial (IA) propuesta en la Comunicación de la Comisión Europea sobre la IA¹:

El término «inteligencia artificial» (IA) se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción —con cierto grado de autonomía— con el fin de alcanzar objetivos específicos.

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).

En este documento ampliamos esta definición para aclarar determinados aspectos de la IA en tanto que disciplina científica y tecnología, con el objetivo de evitar malentendidos, alcanzar una comprensión común de la IA que también puedan utilizar de forma fructífera las personas que no sean expertas en la materia y ofrecer detalles útiles tanto en el debate sobre las directrices éticas de la IA como en las recomendaciones sobre políticas en este campo.

1. Sistemas de inteligencia artificial

El término IA contiene una referencia explícita al concepto de inteligencia. Ahora bien, dado que la inteligencia (sea en las máquinas o en los seres humanos) es un concepto vago pese a haber sido estudiado en profundidad por psicólogos, biólogos y neurocientíficos, los investigadores que trabajan en el terreno de la IA utilizan principalmente el término «racionalidad». La racionalidad hace referencia a 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. Por supuesto, la racionalidad no es el único ingrediente del concepto de inteligencia, pero sí uno muy importante.

En este documento utilizaremos la expresión *sistema de IA* para referirnos a cualquier componente, software o hardware basado en IA. De hecho, los sistemas de IA suelen estar *integrados* en forma de componentes de sistemas más amplios; no es tan frecuente encontrarlos como sistemas autónomos.

Por lo tanto, un sistema de IA es, ante todo, racional, de acuerdo con uno de los manuales de inteligencia artificial más utilizados². Pero ¿cómo alcanza la racionalidad un sistema de IA? Como señala la primera frase de la definición operativa del concepto de IA anteriormente expuesta, lo hace percibiendo el entorno en el que se encuentra inmerso el sistema a través de sensores, recopilando e interpretando datos, razonando sobre lo que percibe o procesando la información derivada de esos datos, decidiendo cuál es la mejor acción que puede realizar y actuando en consecuencia mediante accionadores, pudiendo así modificar el entorno. Los sistemas de IA pueden utilizar normas simbólicas o aprender un modelo numérico; también pueden adaptar su conducta mediante el análisis del modo en que el entorno se ve afectado por sus acciones anteriores. La ilustración 1 muestra un sistema de IA.

¹ Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Inteligencia Artificial para Europa, Bruselas, 25.4.2018 COM(2018) 237 final.

² *Artificial Intelligence: A Modern Approach*, S. Russell y P. Norvig, Prentice Hall, 3.ª edición, 2009.

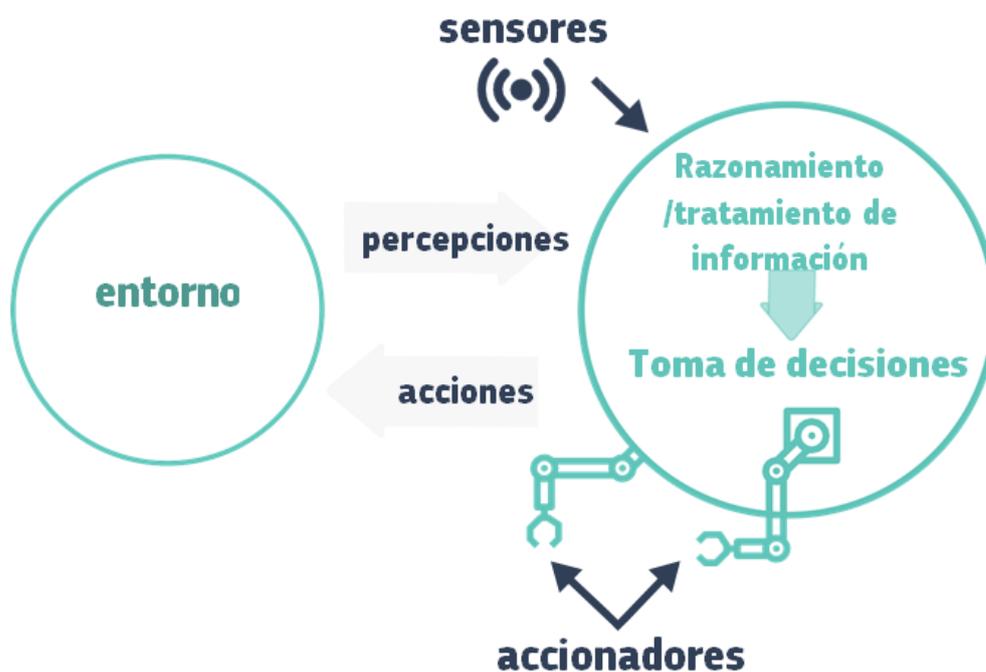


Ilustración 1: Descripción esquemática de un sistema de inteligencia artificial.

Sensores y percepción. En la ilustración 1, los sensores del sistema aparecen reflejados con el símbolo del wifi. En la práctica, puede tratarse de cámaras, micrófonos, un teclado, un sitio web u otros dispositivos de captación de información tales como sensores de cantidades físicas (por ejemplo, temperatura, presión, distancia, fuerza/torsión, sensores táctiles). En general, es necesario dotar al sistema de IA de sensores adecuados para percibir los datos presentes en el entorno que resultan pertinentes para lograr el objetivo establecido para el sistema de IA por su diseñador humano. Por ejemplo, si deseamos construir un sistema de IA que limpie automáticamente el suelo de una habitación cuando esté sucio, los sensores pueden incluir cámaras que tomen imágenes del suelo.

En cuando a los datos recopilados, a menudo resulta útil distinguir entre datos estructurados y no estructurados. Se entiende por *datos estructurados* aquellos que están organizados de acuerdo con modelos predefinidos (como ocurre en el caso de una base de datos relacional); los *datos no estructurados*, en cambio, no tienen una organización conocida (como sucede con una imagen o un fragmento de texto).

Razonamiento, tratamiento de información y toma de decisiones. En el «corazón» de un sistema de inteligencia artificial se encuentra su módulo de razonamiento o tratamiento de información, que recibe datos a través de los sensores, los procesa y propone una acción de acuerdo con el objetivo fijado. Esto significa que los datos recabados por los sensores deben convertirse en información que el módulo de razonamiento o tratamiento de información pueda comprender. Siguiendo con nuestro ejemplo de un sistema de limpieza basado en IA, la cámara proporcionará una imagen del suelo al módulo de razonamiento o tratamiento de información; este deberá decidir si tiene que limpiar el suelo o no (es decir, cuál es la acción óptima para lograr el objetivo deseado). Pese a que para los seres humanos puede resultar sencillo pasar de la imagen del suelo a la decisión de si es necesario limpiarlo o no, para una máquina no es tan fácil, porque una imagen no es más que una secuencia de ceros y unos. Por lo tanto, el módulo de razonamiento o tratamiento de información debe:

1. interpretar la imagen para decidir si el suelo está limpio o no. En general, esto significa que ha de ser capaz de transformar datos en información y procesar esa información de manera sucinta, aunque incluyendo todos los datos pertinentes (en este caso, si el suelo está limpio o no);

2. razonar a partir de ese conocimiento o procesar esa información para elaborar un modelo numérico (esto es, una fórmula matemática) para decidir cuál es la mejor acción. En este ejemplo, si la información derivada de la imagen es que el suelo se encuentra sucio, la acción óptima será iniciar la limpieza; de lo contrario, la mejor acción será permanecer en espera.

Obsérvese que es necesario entender el término «decisión» de forma amplia como cualquier acto conducente a seleccionar la acción que se debe llevar a cabo, y no significa necesariamente que los sistemas de IA sean completamente autónomos. Una decisión también puede consistir en la elección de una recomendación que formular a un ser humano, que será quien tome la decisión final.

Accionamiento. Una vez decidida la acción, el sistema de IA está listo para ejecutarla a través de los accionadores de los que dispone. En la ilustración anterior, los accionadores aparecen representados en forma de brazos articulados, pero no es necesario que sean accionadores físicos. También pueden adoptar la forma de software. En nuestro ejemplo, el sistema de IA podría generar una señal que active un aspirador si el sistema decide llevar a cabo la acción «limpiar el suelo». Otro ejemplo puede ser un sistema conversacional (como un chatbot) que actúa generando textos para responder a las declaraciones del usuario.

Es posible que la acción modifique el entorno, de modo que la próxima vez que el sistema necesite utilizar de nuevo sus sensores quizás perciba información diferente debido a que el entorno se haya visto modificado.

Los sistemas racionales de IA no siempre eligen la acción óptima para lograr su objetivo, por lo que puede decirse que su *racionalidad es limitada* debido a restricciones en los recursos, como el tiempo o la capacidad informática.

LOS sistemas racionales de IA constituyen una versión muy básica de los sistemas de inteligencia artificial. Modifican el entorno, pero no adaptan su comportamiento a lo largo del tiempo para lograr mejor su objetivo. Un *sistema racional con capacidad de aprendizaje* es un sistema racional que, después de llevar a cabo una acción, evalúa el nuevo estado del entorno (a través de la percepción) para determinar el grado de éxito de dicha acción y, en función del resultado, adapta sus reglas de razonamiento y sus métodos de toma de decisiones.

2. La IA como disciplina científica

El sistema de IA anterior constituye una descripción muy simple y abstracta de este tipo de sistemas, con tres capacidades fundamentales: percepción, razonamiento/toma de decisiones y acción. No obstante, es suficiente para permitirnos presentar y comprender la mayoría de las técnicas y subdisciplinas de IA que se utilizan actualmente para desarrollar los sistemas de IA, puesto que todas ellas hacen referencia a las diversas capacidades de estos sistemas. En términos generales, todas estas técnicas se pueden agrupar en dos grandes categorías, referidas a la capacidad de *razonamiento y aprendizaje*. La robótica es otra disciplina muy importante.

Razonamiento y adopción de decisiones. Este grupo de técnicas incluye la representación del conocimiento así como el razonamiento, la planificación, la programación, la búsqueda y la optimización. Estas técnicas permiten llevar a cabo el razonamiento a partir de los datos procedentes de los sensores. Para ello es necesario transformar los datos en conocimiento, de modo que una de las áreas de la inteligencia artificial está relacionada con la optimización de la modelización de ese conocimiento (*representación del conocimiento*). Una vez modelizado el conocimiento, el siguiente paso consiste en razonar con él (*razonamiento a partir del conocimiento*), lo que incluye la extracción de inferencias a través de reglas simbólicas, actividades de *planificación y programación*, la *búsqueda* en un amplio conjunto de soluciones y la *optimización* entre todas las soluciones posibles a un problema. El paso final es la decisión sobre la acción que se debe llevar a cabo. El componente de razonamiento o toma de decisiones de un sistema de IA suele presentar una elevada complejidad y requiere una combinación de varias de las técnicas anteriormente mencionadas.

Aprendizaje. Este grupo de técnicas incluye el aprendizaje automático, las redes neuronales, el aprendizaje profundo, árboles de decisión y muchas otras técnicas de aprendizaje. Estas técnicas permiten a un sistema de IA aprender a resolver problemas que no se pueden especificar de manera precisa, o cuyo método de resolución no se pueda describir mediante reglas de razonamiento simbólicas. Este tipo de problemas tiene que ver, por ejemplo, con

las capacidades de percepción, como el *habla* o la *comprensión del idioma*, así como con la *visión informática* o la *predicción del comportamiento*. Obsérvese que se trata de problemas aparentemente sencillos, porque así suelen ser para las personas. Sin embargo, los sistemas de IA tienen mayores dificultades, ya que (al menos por el momento) carecen de la capacidad de razonar guiándose por el sentido común. La dificultad aumenta cuando el sistema necesita interpretar datos no estructurados. Aquí es donde las técnicas relacionadas con el *aprendizaje automático* pueden resultar de gran ayuda. No obstante, dichas técnicas se pueden utilizar para muchas otras tareas además de la mera percepción. Las técnicas de aprendizaje automático producen un modelo numérico (es decir, una fórmula matemática) que se utiliza para calcular la decisión a partir de los datos disponibles.

Existen varias versiones de aprendizaje automático. Los enfoques más extendidos son el *aprendizaje supervisado*, el *aprendizaje no supervisado* y el *aprendizaje por refuerzo*.

En el caso del aprendizaje automático supervisado, en lugar de proporcionar normas de comportamiento al sistema, introducimos en él ejemplos de comportamiento de entrada y salida, confiando en que será capaz de generalizar a partir de ejemplos (normalmente referidos al pasado) y de comportarse adecuadamente en situaciones no descritas en esos ejemplos (con las que el sistema puede encontrarse en el futuro). En nuestro ejemplo, proporcionaríamos al sistema numerosos ejemplos de imágenes de un suelo y su correspondiente interpretación (esto es, si en cada una de esas imágenes el suelo está limpio o sucio). Si le facilitamos suficientes ejemplos, que sean variados y representativos de la mayoría de las situaciones, el sistema, a través de su algoritmo de aprendizaje automático, será capaz de generalizar para saber cómo debe interpretar correctamente las imágenes de suelos que nunca haya visto antes. Algunos métodos de aprendizaje automático adoptan algoritmos basados en el concepto de *redes neuronales*, que, a grandes rasgos, se inspira en el cerebro humano, en el sentido de que cuenta con una red de unidades de procesamiento de pequeño tamaño (de forma análoga a las neuronas humanas) con numerosas conexiones ponderadas entre ellas. Una red neuronal utiliza los datos que recibe de los sensores (en nuestro ejemplo, la imagen del suelo) y produce una interpretación de la imagen (si el suelo está limpio o no). Durante el análisis de los ejemplos (la fase de *adiestramiento* o formación de la red), las ponderaciones de las conexiones se ajustan para adecuarlas al máximo a la información que proporcionan los ejemplos (con el fin, por tanto, de minimizar la desviación entre el resultado esperado y el resultado calculado por la red). Al término de la fase de formación, se verifica si la tarea se ha aprendido correctamente a través de una fase de comprobación del comportamiento de la red neuronal en los ejemplos con los que nunca se había encontrado anteriormente.

Es importante tener en cuenta que este enfoque (como sucede con todas las técnicas de aprendizaje automático) siempre presenta cierto porcentaje de error, aunque por lo general es reducido. Por lo tanto, un concepto esencial es el de *precisión*, que mide la proporción de respuestas correctas.

Existen varios tipos de redes neuronales y de métodos de aprendizaje automático; uno de los que están cosechando mayores éxitos en la actualidad es el *aprendizaje profundo*. Este enfoque se refiere al hecho de que la red neuronal cuenta con varias capas entre la entrada y la salida, que le permiten aprender la relación general entre ellas en pasos sucesivos. De este modo, este enfoque resulta más preciso y requiere un menor grado de orientación humana.

Las redes neuronales son solamente una herramienta de aprendizaje automático, pero existen muchas otras con diferentes propiedades: bosques aleatorios y árboles potenciados, métodos de agrupación, factorización matricial, etc.

Otro método muy útil de aprendizaje automático se denomina *aprendizaje por refuerzo*. En este caso se deja libertad al sistema de IA para tomar sus decisiones a lo largo del tiempo, y cada vez que adopta una decisión le proporcionamos una señal de recompensa que informa al sistema si la decisión fue acertada o no. El objetivo del sistema es maximizar a lo largo del tiempo las recompensas positivas que recibe. Este enfoque se utiliza, por ejemplo, en sistemas de formulación de recomendaciones (como en el caso de diversos sistemas de este tipo disponibles en internet que sugieren a los usuarios productos que quizá les gustaría adquirir), así como en el mundo del marketing.

Los métodos de aprendizaje automático no solo son útiles en el campo de las tareas de percepción, sino también en el ámbito de la visión y la comprensión de textos; no obstante, en todas estas áreas resulta complicado definir las tareas y no es posible definir las de forma exhaustiva mediante reglas simbólicas de comportamiento.

Obsérvese la distinción entre los enfoques de aprendizaje automático dirigidos a aprender una tarea nueva que no es posible describir adecuadamente de manera simbólica y los agentes racionales con capacidad de aprendizaje (mencionados en la sección anterior), que adaptan su comportamiento a lo largo del tiempo para mejorar la consecución de su objetivo. Estas dos técnicas pueden solaparse o cooperar entre ellas, pero no son necesariamente idénticas.

Robótica. La robótica se puede definir como una «acción de la IA en el mundo físico» (también denominada *IA integrada*). Un robot es una máquina física que debe enfrentarse a la dinámica, la incertidumbre y la complejidad del mundo físico. Normalmente, la percepción, el razonamiento, la acción, el aprendizaje y las capacidades de interacción con otros sistemas suelen integrarse en la arquitectura del control del sistema robótico. Además de la IA, existen otras disciplinas que desempeñan algún tipo de función en el diseño y funcionamiento de robots, como la ingeniería mecánica o la teoría del control. Los robots incluyen, por ejemplo, manipuladores robóticos, vehículos autónomos (como coches, drones, taxis voladores, etc.), robots humanoides, robots de limpieza, etc.

La ilustración 2 muestra la mayoría de las subdisciplinas mencionadas anteriormente, así como la relación que existe entre ellas. Es importante, sin embargo, tener en cuenta que la IA es un campo mucho más complejo de lo que refleja esta ilustración, puesto que incluye muchas otras subdisciplinas y técnicas. Además, como ya se ha indicado, la robótica también se apoya en técnicas ajenas al espacio de la IA. Entendemos, no obstante, que esta representación es suficiente para mostrar de manera adecuada el tipo de intercambio de opiniones, concienciación y debate sobre IA, ética de la IA y políticas relativas a la IA que es necesario acometer en el seno del propio grupo de expertos de alto nivel, un grupo multidisciplinar formado por múltiples partes interesadas.



Ilustración 2: Descripción general simplificada de las subdisciplinas de la IA y de la relación existente entre ellas.

Tanto el aprendizaje automático como el razonamiento incluyen muchas otras técnicas; la robótica, por su parte, incluye técnicas ajenas a la IA. La inteligencia artificial es un subconjunto de la informática.

3. Otros conceptos y cuestiones importantes relacionados con la inteligencia artificial

IA específica (o débil) y general (o fuerte). Un sistema de IA general es un sistema diseñado para realizar la mayoría de las actividades que pueden llevar a cabo los seres humanos. Por el contrario, los sistemas de IA específicos solamente pueden realizar una tarea concreta o un número reducido de ellas. Los sistemas de IA desplegados actualmente constituyen ejemplos de IA específica. En las etapas iniciales de la inteligencia artificial, los

investigadores utilizaban una terminología diferente (IA «débil» y «fuerte»). Hoy en día continúan existiendo numerosos desafíos éticos, científicos y tecnológicos sin resolver para desarrollar las capacidades que serán necesarias para construir sistemas de IA generales, como el razonamiento basado en el sentido común, la autoconciencia o la capacidad de la máquina para definir su propio objetivo.

Problemas y sesgos de los datos. Dado que muchos sistemas de IA, como los que incluyen componentes de aprendizaje automático supervisado, se apoyan en enormes cantidades de datos para realizar correctamente su función, es importante comprender cómo influyen esos datos en el comportamiento de los sistemas de IA. Por ejemplo, si los datos de formación presentan sesgos —es decir, si no son suficientemente equilibrados o inclusivos—, el sistema de IA formado a partir de esos datos no podrá extraer generalizaciones correctas y existe la posibilidad de que tome decisiones injustas que favorezcan a unos grupos sobre otros. En fechas recientes, la comunidad de IA ha trabajado en el desarrollo de métodos para detectar y mitigar los sesgos de los conjuntos de datos de formación, así como en otras partes de un sistema de IA.

IA de caja negra y explicabilidad. Algunas técnicas de aprendizaje automático, pese a haber cosechado un éxito muy importante desde el punto de vista de la precisión, resultan altamente opacas, lo que dificulta la comprensión del modo en que adoptan sus decisiones. El concepto de *IA de caja negra* se refiere a este tipo de escenarios en los que no es posible encontrar el motivo subyacente a determinadas decisiones. La explicabilidad es una propiedad de los sistemas de IA que, por el contrario, pueden explicar de algún modo sus acciones.

IA orientada al logro de objetivos. Los sistemas actuales de inteligencia artificial están orientados al logro de objetivos; esto significa que reciben por parte de un ser humano la especificación del objetivo que deben conseguir y utilizan determinadas técnicas para alcanzarlo. Por lo tanto, no definen sus propios objetivos. No obstante, algunos sistemas de IA (como los basados en ciertas técnicas de aprendizaje automático) pueden tener mayor libertad para decidir el curso de actuación a fin de alcanzar el objetivo fijado.

4. Definición actualizada de inteligencia artificial

Proponemos utilizar la siguiente definición actualizada del concepto de inteligencia artificial:

«Los sistemas de inteligencia artificial (IA) son sistemas de software (y en algunos casos también de hardware) diseñados por seres humanos³ que, dado un objetivo complejo, actúan en la dimensión física o digital mediante la percepción de su entorno a través de la obtención de datos, la interpretación de los datos estructurados o no estructurados que recopilan, el razonamiento sobre el conocimiento o el procesamiento de la información derivados de esos datos, y decidiendo la acción o acciones óptimas que deben llevar a cabo para lograr el objetivo establecido. Los sistemas de IA pueden utilizar normas simbólicas o aprender un modelo numérico; también pueden adaptar su conducta mediante el análisis del modo en que el entorno se ve afectado por sus acciones anteriores.

La IA es una disciplina científica que incluye varios enfoques y técnicas, como el aprendizaje automático (del que el aprendizaje profundo y el aprendizaje por refuerzo constituyen algunos ejemplos), el razonamiento automático (que incluye la planificación, programación, representación y razonamiento de conocimientos, búsqueda y optimización) y la robótica (que incluye el control, la percepción, sensores y accionadores así como la integración de todas las demás técnicas en sistemas ciberfísicos)»,

y hacer referencia a este documento como fuente de información adicional para respaldar esta definición.

³ Los seres humanos diseñan sistemas de IA directamente, aunque también pueden emplear técnicas de IA para optimizar su diseño.

**Este documento ha sido elaborado por los miembros del Grupo de expertos de alto nivel
sobre inteligencia artificial,**

que se relacionan a continuación en orden alfabético:

Pekka Ala-Pietilä, presidente del Grupo de expertos de alto nivel AI Finland, Huhtamaki, Sanoma	Pierre Lucas Orgalim – Industrias tecnológicas europeas
Wilhelm Bauer Fraunhofer	Ieva Martinkenaite Telenor
Urs Bergmann Zalando	Thomas Metzinger JGU Mainz y Asociación Europea de Universidades
Mária Bieliková Universidad eslovaca de Tecnología, Bratislava	Cateljine Muller ALLAI Netherlands y Comité Económico y Social Europeo (CESE)
Cecilia Bonefeld-Dahl DigitalEurope	Markus Noga SAP
Yann Bonnet ANSSI	Barry O’Sullivan, vicepresidente del Grupo de expertos de alto nivel Colegio Universitario de Cork
Loubna Bouarfa OKRA	Ursula Pahl BEUC
Stéphan Brunessaux Airbus	Nicolas Petit Universidad de Lieja
Raja Chatila Iniciativa del IEEE para la Ética de los Sistemas Inteligentes/Autónomos y Universidad de la Sorbona	Christoph Peylo Bosch
Mark Coeckelbergh Universidad de Viena	Iris Plöger BDI
Virginia Dignum Universidad de Umeå	Stefano Quintarelli Garden Ventures
Luciano Floridi Universidad de Oxford	Andrea Renda College of Europe Faculty y CEPS
Jean-Francois Gagné Element AI	Francesca Rossi* IBM
Chiara Giovannini ANEC	Cristina San José Federación Bancaria Europea
Joanna Goodey Agencia de los Derechos Fundamentales	George Sharkov Digital SME Alliance
Sami Haddadin Escuela de Robótica de Múnich y MI	Philipp Slusallek Centro Alemán de Investigación en IA (DFKI)
Gry Hasselbalch Laboratorio de ideas DataEthics y Universidad de Copenhague	Françoise Soulié Fogelman Consultora de IA
Fredrik Heintz Universidad de Linköping	Saskia Steinacker Bayer
Fanny Hidvegi Access Now	Jaán Tallinn Ambient Sound Investment
Eric Hilgendorf Universidad de Wurzburg	Thierry Tingaud STMicroelectronics
Klaus Höckner Hilfsgemeinschaft der Blinden und Sehschwachen	Jakob Uszkoreit Google
Mari-Noëlle Jégo-Laveissière Orange	Aimee Van Wynsberghe TU Delft
Leo Kärkkäinen Nokia Bell Labs	Thiébaud Weber Confederación Europea de Sindicatos (CES)
Sabine Theresia Köszegi TU Wien	Cecile Wendling AXA
Robert Kroplewski Abogado y asesor del Gobierno de Polonia	Karen Yeung Universidad de Birmingham
Elisabeth Ling RELX	

*Francesca Rossi actuó como ponente de este documento.