Inteligencia artificial: una aproximación jurídica no catastrofista

ALEJANDRO HUERGO LORA
Catedrático de Derecho Administrativo. Universidad de Oviedo

RESUMEN

La inteligencia artificial (IA) es la principal tecnología en desarrollo en estos años y se encuentra, o va a encontrarse, presente en casi todas las ramas de actividad. Por un lado, se esperan de ella grandes avances, básicamente al permitir tomar mejores decisiones y aprovechar el potencial informativo que ofrecen los datos. Ya ha contribuido a crear gigantes empresariales. Por otro lado, tiene riesgos que es necesario identificar. El trabajo realiza una aproximación jurídica partiendo de una definición y clasificación. A continuación se identifican los avances potenciales y también los riesgos. Se esboza un intento de encaje jurídico de la inteligencia artificial. Finalmente se lleva a cabo una aproximación al proyecto de Reglamento de Inteligencia Artificial de la UE, principal avance legislativo en la materia.

PALABRAS CLAVE

Inteligencia artificial | AI Act | Discriminación | Datos

ABSTRACT

Artificial intelligence (AI) is the main technology under development in recent years and is, or will be, present in almost all branches of activity. On the one hand, great advances are expected from it, basically by allowing better decisions to be made and taking advantage of the information potential offered by big data. It has already helped create corporate giants. On the other hand, it has risks that need to be identified. The paper makes a legal approach based on a definition and classification. Potential advances as well as risks are identified. An attempt to legally understand artificial intelligence is outlined. Finally, an approximation to the EU Artificial Intelligence Regulation project is carried out, the main legislative advance in the matter.

KEYWORDS

Artificial intelligence | *AI Act* | *Discrimination* | *Data protection*

Trabajo realizado en el marco del proyecto de investigación PID2021-126881OB-I00 (Herramientas algorítmicas para ciudadanos y administraciones públicas), financiado por el Ministerio de Ciencia e Innovación. Entregado en agosto de 2023.

1. Introducción

La inteligencia artificial (IA) es el gran tema de los últimos años y parece que tiene por delante mucho recorrido. Su presencia en los medios de comunicación es abrumadora. Aparece ligada a casi cualquier descubrimiento científico reciente. Las empresas de mayor capitalización bursátil del mundo (las grandes tecnológicas) tienen en común haber impulsado y explotar comercialmente la IA. En todos los sectores, de la energía a las finanzas, las empresas se esfuerzan en aprovechar esta nueva tecnología para obtener un extra de eficiencia y de rentabilidad.

Con todo, en sólo dos años el panorama ha cambiado sustancialmente. Lo que hace poco tiempo era, ante todo, una promesa de la que se esperaban ganancias no sólo para las empresas sino también para los ciudadanos y los servicios públicos, aparece ahora como una amenaza, una de las mayores amenazas, de hecho, para la humanidad. La aparición estelar de Chat-GPT en la primavera de 2023 dio lugar a una catarata de manifiestos catastrofistas, en alguno de los cuales se pidió una moratoria de 6 meses en la investigación de IA, iniciativa pronto abandonada.

Las amenazas serían de diversos tipos. Al viejo argumento ludita (la IA aniquilaría muchos puestos de trabajo, arrojando a millones de personas al desempleo y la frustración) se une el problema del «engaño»: la IA (en su versión generativa) sería capaz de crear textos, audios y vídeos que no son distinguibles de los reales, lo que engañaría a los ciudadanos y podría convertirse en un instrumento de desinformación y de difusión de bulos. Del mismo modo, la posibilidad de crear textos e imágenes mediante instrumentos de IA generativa distorsionaría algunas actividades (por ejemplo, los estudiantes podrían burlar a sus profesores presentado trabajos elaborados con IA, se podrían presentar como obras de arte originales productos salidos de aplicaciones de IA) y provocaría conflictos jurídicos y situaciones de explotación de la obra ajena (una aplicación de IA, ¿puede escanear indiscriminadamente obras escritas y audiovisuales que se encuentran en la red y utilizarlas para crear «obras» que va a vender, directa o indirectamente?). A ello se une la ya antigua amenaza de los «sesgos» (magistralmente denunciada ya por C. O'Neil en 2016), es decir, la posibilidad de que las predicciones generadas por la IA, determinantes de la actuación de empresas o incluso Administraciones Públicas, sean discriminatorias o perjudiquen a los sujetos más vulnerables[1].

La constatación de estos riesgos lleva a pedir la regulación de la IA para conseguir, como ha ocurrido a lo largo de los años con tantas otras tecnologías de riesgo (toda tecnología provoca riesgos), limitar el peligro y proteger los bienes jurídicos afectados. En eso estamos actualmente. Tras múltiples ensayos y tentativas de carácter no vinculante, el Reglamento europeo sobre IA (denominado, «comercialmente», «Ley de IA») constituye sin duda la punta de lanza.

El Estado (mejor dicho, las instituciones públicas en general, incluida la UE) se encuentra llamado a regular la IA por las razones indicadas, pero a la vez intenta aprovecharla (como ya han hecho las empresas) para obtener de ella beneficios en su actividad. En cierto modo, los Estados están obligados a utilizar la IA para compensar su debilidad y escasez de recur-

1. Me remito, para mayores referencias, al libro dirigido por mí y coordinado por G. Díaz González, *La regulación de los algoritmos* (Cizur Menor, 2020), que contiene múltiples aproximaciones sectoriales, y también a mis trabajos «Administraciones Públicas e inteligencia artificial: ¿más o menos discrecionalidad?», *El Cronista,* 96-97 (2021), págs. 78-95, y «Gobernar con algoritmos, gobernar los algoritmos», *El Cronista,* 100 (2022), págs. 80-89.

sos y para intentar ponerse, de algún modo, al nivel de unas empresas que ya cuentan con el «turbo» que les proporciona esta tecnología. De ahí la dualidad que he utilizado como título de un trabajo anterior: gobernar los algoritmos, gobernar con algoritmos.

2. De qué se habla cuando se habla de inteligencia artificial

2.1. Introducción

Tomando prestado el título de uno de los más asequibles libros de Murakami, me parece imprescindible empezar con una mínima descripción del fenómeno analizado, cuya omnipresencia informativa no va acompañada de explicaciones claras sobre sus contornos. Cabría tomar prestada otra expresión, ésta de Ortega y Gasset, y decir que es muy frecuente «pensar en hueco», es decir, hablar sobre la IA y predicar de ella bondades y riesgos, e incluso postular su regulación, sin partir de una idea mínimamente precisa acerca de qué es y qué puede esperarse de ella[2].

La IA es un ideal casi eterno, como el de volar. Es del todo lógico que la tendencia humana a inventar nuevas técnicas e instrumentos no se detenga e intente replicar también las funciones más específicamente humanas, las de tipo intelectual, cuya reproducción constituiría el reto máximo a que se enfrenta la mente humana. De todas formas, conviene decir que estamos bastante lejos de ese objetivo, sin perjuicio de que los avances ya realizados en tantos campos obliguen a ser muy prudentes a la hora de descartar cualquier tipo de logro tecnológico en el futuro, ante el grave de riesgo de hacer el ridículo si se trata de poner puertas al campo. Como también he dicho en otro lugar, el ideal de la IA es tan legítimo como el de volar, pero lo que ahora tenemos se parece a ese ideal tanto como la aviación comercial a las alas que imaginaba Leonardo da Vinci.

Por otro lado, la IA es un antropomorfismo, una metáfora. Los sistemas de IA producen el mismo resultado que los operadores humanos, en el sentido de que hacen cosas que, si no se hacen con IA, exigen la actuación de un operador humano[3]. Sin embargo, la forma de lograr tal objetivo es distinta. Los sistemas de IA no «piensan», sino que aplican «fórmulas» matemáticas complejas, creadas por humanos a partir de grandes cantidades de datos.

---

2. «Una misma cosa se puede pensar de dos modos: en hueco o en lleno. Si decimos que la historia se propone averiguar cómo han sido las vidas humanas, se puede estar seguro que el que nos escucha al entender estas palabras y repetírselas las piensa en hueco, esto es, no se hace presente la realidad misma que es la vida humana, no piensa, pues, efectivamente el contenido de esa idea, sino que usa aquellas palabras como un continente vacío, como una ampolla inane que lleva por de fuera el rótulo: «vida humana». Es, pues, como si se dijera: Bueno, yo me doy cuenta de que al pensar ahora estas palabras —al leerlas, oírlas o pronunciarlas— no tengo de verdad presente la cosa que ellas significan, pero tengo la creencia, la confianza de que siempre que quiera detenerme a realizar su significado, a hacerme presente la realidad que nombran, lo conseguiré. Las uso, pues, fiduciariamente, a crédito, como uso un cheque, confiado en que siempre que quiera lo podré cambiar en la ventanilla de un Banco por el dinero contante y sonante que representa. Confieso que, en rigor, no pienso mi idea, sino sólo su alvéolo, su cápsula, su hueco. Este pensar en hueco y a crédito, este pensar algo sin pensarlo en efecto es el modo más frecuente de nuestro pensamiento» (*En torno a Galileo,* 1933, lección III: «La idea de la generación»).

3. Existe una cierta indefinición lingüística a la hora de referirse a los sistemas de IA. Ésta (la inteligencia artificial) es un fenómeno muy amplio, un término genérico. Lo que funciona en la práctica son «sistemas de IA» que hacen cosas concretas (esa es la expresión que utiliza el Reglamento, «sistemas de IA»). Podríamos hablar también de «aplicaciones». En el lenguaje periodístico se utiliza con frecuencia la expresión IA en sentido concreto («contable», como diríamos en inglés) para referirse a sistemas de IA. Por ejemplo, cuando se dice que «una inteligencia artificial predice los resultados de los partidos de baloncesto».

2.2. De la automatización a la IA

Es necesario distinguir la IA de la mera automatización de procesos o decisiones. La automatización o robotización consiste en que las «máquinas» sustituyan a los humanos en la realización de una tarea, para lo que se necesita describir y estandarizar la tarea de forma que pueda ser reproducida por una máquina (una máquina dirigida, cuando la tarea tiene una cierta complejidad, por un ordenador). Del mismo modo que se automatiza la cadena de montaje de una fábrica de automóviles, hasta el punto de que puede reducirse al mínimo el número de trabajadores, se puede automatizar la elaboración de ciertas liquidaciones tributarias o de propuestas de resolución en procedimientos sancionadores (a partir, por ejemplo, de las fotos que toma un radar, que se dispara automáticamente cuando detecta el paso de un vehículo a velocidad superior a la máxima permitida).

De la misma forma que la cadena de montaje se programa y se le dan unas instrucciones precisas (lo que exige un previo análisis riguroso de la actividad que llevaban a cabo los operadores humanos), en la automatización de tareas de tipo jurídico se vuelca la norma aplicable al caso en un programa informático, de modo que éste, al ejecutarse, aplica la norma al supuesto de hecho que previamente se le facilita (supuesto de hecho que a veces se obtiene también automáticamente, como en el ejemplo del radar), produciendo como resultado un texto que puede convertirse, bien tras la «firma» del humano correspondiente, o bien de forma automática (cuando ello esté previsto), en una decisión administrativa (y, eventualmente, judicial, puesto que las diferencias entre un caso y otro serán jurídicas, pero no técnicas).

Característica básica de esa automatización es que nos hallemos ante decisiones totalmente regladas, programadas de forma estricta por las normas jurídicas, de tal modo que, a partir de un determinado hecho, la norma imponga una única solución. Si esto es así, no resulta difícil crear un programa que automatice la aplicación de la norma, obteniendo así ganancias considerables, puesto que se podrán dictar muchos más actos en menos tiempo (reduciendo el tiempo de tramitación) e incluso se reducirá el riesgo de errores o de aplicaciones desviadas de la norma. Ojalá se dispusiera de aplicaciones de este tipo, por ejemplo, para reconocer el derecho a la percepción del ingreso mínimo vital (IMV), que, tres años después de su aprobación, sólo ha llegado a una pequeña parte de sus destinatarios potenciales[4]. Sus aplicaciones son ya muy conocidas, por ejemplo, en el ámbito tributario, puesto que prácticamente todos los contribuyentes elaboran su declaración del IRPF a partir de una aplicación informática (ahora denominada Renta Web) que calcula la cuota del impuesto a partir de los datos introducidos por el contribuyente, después de que los programadores hayan «volcado» en ella la legislación aplicable al impuesto.

La automatización no influye en el contenido de las decisiones administrativas y, sobre todo, es irrelevante para determinar si son, o no, conformes al ordenamiento jurídico. La decisión de que se trate será conforme a derecho en la medida en que se ajuste a la legislación aplicable, y, en caso contrario, no lo será. Es irrelevante que se haya elaborado con el

4. Aquí podrían entrar en juego, también, cambios de mentalidad o de «cultura administrativa» que sin duda deberán ensayarse y en los que se hizo algún avance durante la pandemia. Así, pasar de una cultura de teórico «riesgo cero» (en la que sólo se reconocen derechos a quienes han demostrado exhaustivamente el cumplimiento de todos los requisitos, lo que supuestamente limita el riesgo de fraude, pero a cambio garantiza que se producirá una demora que puede equivaler a un incumplimiento de la norma) a una cultura de gestión o distribución de riesgos, en la que se asuma un riesgo limitado de incumplimiento de requisitos (compensado con controles posteriores y la amenaza de sanciones) a cambio de una mayor eficacia.

uso de una aplicación informática o no. De hecho, en la mayoría de ocasiones el destinatario no conoce si se ha utilizado. Por ejemplo, en el programa de subvenciones del «kit digital» se ha hecho un uso intenso de la robotización administrativa y en muchos programas otros no. Para los destinatarios de la subvención es algo indiferente.

Además, para determinar si la medida administrativa generada mediante aplicaciones informáticas es conforme a derecho o no, es innecesario conocer el funcionamiento del sistema informático. Se puede averiguar, «con lápiz y papel», qué decisión tenía que dictar la Administración (por ejemplo, calculando la liquidación del impuesto sin utilizar el sistema informático utilizado por la Administración) y comprobar de ese modo si la decisión realmente tomada era correcta o no. Esto quita importancia a los problemas de transparencia y de «acceso al código fuente», porque no es necesario para saber si la decisión es correcta o no.

Por supuesto, la automatización, a la vez que incrementa la capacidad productiva de la Administración (el mismo número de funcionarios elaboran muchos más actos administrativos en menos tiempo, lo que resulta, en principio, muy beneficioso para los ciudadanos y para los intereses públicos) y reduce el riesgo de que se cometan errores o de que individuos concretos se desvíen de la norma, también hace posible que, si se cometen errores u omisiones en la elaboración del programa, estos errores afecten a miles de actos administrativos y sean difíciles de corregir. Todos conocemos lo difícil que es «forzar» a una aplicación informática o intentar razonar con un operador humano que no puede saltarse los parámetros de la aplicación que constituye su instrumento de trabajo, aunque él mismo o ella misma se den cuenta del error. Todo esto justifica que se tomen precauciones ante la puesta en práctica de este tipo de aplicaciones y que se establezcan mecanismos de reacción ante posibles errores. Jurídicamente, se encuentran en el artículo 22 RGPD, el artículo 41 de la Ley 40/2015 y en normas y proyectos autonómicos.

Pero ello no debe ocultar que, frente a lo que veremos que es característico de la IA (que es que influye en el contenido de la actividad administrativa, puesto que éste se derivará de lo que dice el sistema de IA, no de lo que dice una norma), aquí la aplicación informática es un fiel aplicador de la norma. De ahí se deriva que este tipo de automatización sólo puede funcionar en contextos de administración reglada, es decir, en los que la norma determina completamente la actuación administrativa.

Si existe un margen decisional, la robotización administrativa no tiene sentido o se convierte en otra cosa. Puede ocurrir que la aplicación informática sirva para aplicar la norma y le deje al operador humano tomar la parte final de la decisión, en la que se introduce ese margen decisional. Por ejemplo, una aplicación que sirva para determinar qué solicitantes reúnen los requisitos para participar en un procedimiento selectivo de acceso a la función pública y le entregue la lista al tribunal para que éste corrija las pruebas, aplicando, allí donde sea procedente, la discrecionalidad técnica. El supuesto es igual al anterior, sólo que la robotización se aplica únicamente a la parte reglada de la decisión.

Pero, salvo este supuesto, la robotización es incompatible con la existencia de márgenes decisionales no programados por la norma. Nuestro Derecho administrativo controla al poder público identificando los momentos en que éste es ejercido y articulando unos controles sobre el mismo. El principio de legalidad supone una programación normativa de la actividad administrativa a través de Leyes y reglamentos, sometidos a control. El acto administrativo también está sometido a control, en el que se distingue, lógicamente, aquellas

partes del acto que son estricta aplicación de las normas y aquellas otras que, ante la ausencia de un parámetro normativo estricto, son fruto de la decisión del órgano que dicta el acto. E incluso existen instrumentos intermedios, como las instrucciones o circulares, que no son normas jurídicas pero influyen en el contenido del acto, y que cada ordenamiento decide si son objeto de control directo o sólo pueden ser atacadas indirectamente al hilo de su aplicación al casos concretos. Pero, en definitiva, las decisiones no pueden escamotearse.

Si aplicamos estas ideas a la automatización o robotización administrativa, llegamos a la conclusión de que no se pueden (o, en todo caso, no se deben) introducir en el programa informático parámetros no previstos en la norma y que predeterminen el contenido de los actos administrativos (estrechándolo). Y ello porque no se ha atribuido potestad alguna para efectuar esta concreción. Jurídicamente, existen dos momentos: la norma y el acto. Si se introducen esos parámetros o restricciones (lo que a veces es fruto de un error, de que el programa es defectuoso y no toma en consideración datos que la norma sí contempla), su control tendrá que hacerse en el acto administrativo.

Pongamos un ejemplo: un programa informático elaborado para calcular el número de horas de clase que debe impartir cada profesor de una Universidad. Se trata de una decisión reglada, puesto que ese número es el resultado de la legislación y de un reglamento de la Universidad. El programa es una calculadora, no introduce ningún parámetro nuevo y no puede hacerlo. Si, por error, el programa distorsiona la aplicación de la norma (por ejemplo, no tiene en cuenta el número de sexenios, aunque la normativa dice que influyen en el número de horas de clase), los actos de aplicación serán ilegales y tendrán que ser anulados en vía de recurso. Para ello no es necesario identificar en qué línea del código fuente se produjo el error, simplemente bastará constatar la contradicción entre el acto y la norma.

Pensemos en otro ejemplo en el que la Administración tiene un margen de decisión: la graduación de sanciones administrativas dentro de la horquilla legalmente establecida. La legislación establece unos parámetros de graduación (artículo 29.3 de la Ley 40/2015). Los jueces pueden corregir la aplicación de estos criterios que se hace en la resolución sancionadora cuando tal aplicación es errónea o claramente desproporcionada, aunque suelen reconocer un «margen de apreciación» a la Administración. ¿Se puede elaborar un programa informático que concrete la sanción —calculando, por ejemplo, el importe de la multa— a partir de una serie de reglas que supongan una concreta forma —no la única posible— de aplicar los criterios legales? Es pertinente recordar la sentencia que, hace años, declaró nula una ordenanza local que pretendía concretar el importe de las multas de tráfico y aumentar la seguridad jurídica, fijando para cada caso un importe concreto dentro de la horquilla establecida en la Ley. La sentencia consideró que la Ley imponía lo que podemos denominar, en terminología alemana, «reserva de acto administrativo»[5]. Creo que se pueden sostener dos afirmaciones.

5. Sentencia de la Sala de lo Contencioso-Administrativo del TSJ de Madrid de 18 de junio de 2009 (recurso 8/2006): «Del examen de la disposición impugnada se desprende claramente que se infringe el principio de jerarquía normativa, pues establece una cuantía fija para cada tipo de infracción sin tener en cuenta que la ley establece un margen o recorrido de la sanción, a modo de ejemplo para las infracciones leves serán sancionadas con multa de hasta 90 euros, lo que implica que pueden ser sancionadas con una multa inferior a los 90 euros que de modo general establece la norma, que es en el concreto acto de imposición de la sanción cuando de manera motivada se debe elegir la concreta sanción a imponer teniendo en consideración todos los elementos del caso, y no puede de manera apriorística determinarse el importe de la sanción».

La primera es que un ordenador no es un título habilitante de potestades administrativas. Si alguien no tiene la potestad de aprobar una norma reglamentaria sobre una determinada cuestión (norma que serviría, obviamente, para dirigir en una determinada dirección la actuación de los órganos administrativos), no puede hacerlo con un programa informático que también predetermine la actuación administrativa, orientándola hacia una opción concreta de las permitidas por la norma. Y, por la misma razón, si alguien tiene esa potestad reglamentaria deberá ejercerla en la forma prevista (mediante una norma publicada en el boletín oficial) y después de tramitar el procedimiento legalmente establecido, no elaborando un programa para que lo utilice la Administración. Otra cosa es que, aprobada la norma, se elabore un programa para facilitar su aplicación. Pero no cabe confundir esos dos objetos.

La segunda afirmación es que esa posible predeterminación del contenido del acto administrativo a través de un programa informático es jurídicamente irrelevante. Pensemos en la sanción cuyo importe concreto deriva de una aplicación informática, que, de todas las posibilidades incluidas en la horquilla legal, se queda con una concreta. Eso no vale como justificación. El acto administrativo deberá motivar ese importe, y dicha motivación tendrá que revisarse o examinarse por sí misma, sin que sirva como justificación (para salvar la validez del acto administrativo) el hecho de apoyarse en una aplicación informática. Puede suceder, incluso, que el acto se anule porque el juez considere que la motivación es ficticia o aparente, dirigida únicamente a justificar de manera formularia una decisión ya adoptada y basada en una razón carente de justificación (el seguimiento ciego de un programa informático).

2.3. Automatización en contextos administrativos complejos

Un paso más se da cuando se utilizan aplicaciones informáticas en contextos en los que, aunque no existe margen decisional alguno para el operador (por lo que seguimos en el campo de la actuación administrativa reglada), tampoco es posible comprobar «manualmente» (es decir, ignorando la existencia de la aplicación informática) si la decisión administrativa es correcta o no.

Comencemos con un ejemplo aparentemente banal, pero que no lo es tanto (y que ha aparecido en decisiones de los órganos de transparencia)[6]. En algunos procedimientos administrativos se prevé la celebración de sorteos, por ejemplo, para determinar todos o alguno de los miembros de tribunales de selección o el orden de actuación de los solicitantes o la letra por la que se empieza en la asignación de bienes escasos por orden alfabético. En muchos casos, los sorteos con bombo y bolas han sido sustituidos por aplicaciones informáticas. Pues bien, del mismo modo que los afectados pueden querer inspeccionar las bolas y el bombo de un sorteo, también pueden aspirar a conocer el funcionamiento del algoritmo que determina el resultado del sorteo informático. Es imposible saber, sin conocer ese programa, si su resultado es realmente aleatorio, por lo que, a diferencia de lo que sucedía en el caso anterior, la aplicación informática no puede ser invisible ni irrelevante.

Algo parecido sucede cuando, como ocurre con frecuencia, se utilizan aplicaciones informáticas para gestionar procedimientos complejos de reparto de bienes escasos que afectan a amplios grupos de interesados, frecuentemente en materia de personal. Por ejemplo, con-

6. Me remito a mi reciente trabajo «El derecho de transparencia en el acceso a los códigos fuente», en *Anuario de Transparencia Local,* 2023, páginas 35-66.

cursos de traslados masivos de personal docente (como ha sucedido en Italia) o la asignación de plazas MIR. La utilización de ayudas informáticas permite agilizar el procedimiento, aunque con frecuencia despierta resistencias porque hay menos transparencia y los interesados se ven obligados a confiar en un mecanismo que no conocen. Es algo parecido a lo que sucede con el voto electrónico en los procesos electorales, que tendría muchas ventajas (inmediatez del escrutinio, posibilidad de votar a distancia sin la incomodidad del voto por correo, reducción del personal necesario para las mesas electorales), pero que probablemente no se ha implantado en países como España porque supone perder la actual transparencia del proceso, basada en factores tan claros como el secreto garantizado por el voto en urna física.

Se trata de procesos completamente reglados porque las decisiones están sometidas a normas (de baremación de méritos y/o de establecimiento de reglas claras de prelación) que no dejan margen alguno a la decisión del operador, por lo que se trata de procesos muy adecuados para su automatización.

Desde el punto de vista jurídico, lo más importante es que, como hemos visto en el párrafo anterior, es imposible verificar la corrección del proceso sin acceder al funcionamiento de la aplicación informática, puesto que, como cada paso (cada asignación de plaza) está condicionado por los anteriores y condiciona, a su vez, a los siguientes, no es posible verificar individualmente si se ajusta a las normas aplicables.

2.4. Inteligencia artificial

Cuando hablamos de inteligencia artificial, se hace referencia a un fenómeno bastante distinto de la automatización o robotización que se acaba de analizar. Se ha dicho, gráficamente, que una cadena montaje de una fábrica de automóviles, en la que cada robot sigue unas instrucciones fijas porque se asume que las piezas le van a ser colocadas de una forma precisa, es un ejemplo claro de automatización, mientras que una aspiradora doméstica automática (o su equivalente que siega el césped en pequeñas parcelas) es un ejemplo de inteligencia artificial[7]. El segundo, a diferencia del primero, puede encontrarse con obstáculos (una silla, un juguete, las patas de cualquier mueble, el borde de una piscina) y no cuenta con un plano definido de las habitaciones que va a aspirar. Este tipo de aparatos tienen que ser capaces de identificar un obstáculo (sin necesidad de que previamente se les indique exactamente qué es un obstáculo y dónde va a encontrarse) y reaccionar ante él (desviándose). Necesitan, por tanto, una cierta capacidad de adaptarse a lo desconocido o de deducir conceptos generales (un obstáculo) a partir de experiencias previas de las que se les haya instruido.

En definitiva, un sistema de IA debe servir para resolver una pregunta, hacer una predicción o tomar una decisión. Y el paso para crearlo es tomar una gran cantidad de datos sobre casos pasados en los que se haya planteado esa cuestión. Normalmente, son muchos más datos de los que puede analizar una sola persona a partir de su experiencia, puesto que la capacidad de almacenamiento y de cálculo de los ordenadores es superior. Se analizan esos datos por ordenadores que, aun siendo «ciegos» (el ordenador no siente el calor ni el frío, pero puede hacer cálculos sobre temperatura; no puede apreciar un buen vino, pero puede analizar miles de opiniones sobre vinos emitidas en redes sociales), pueden hallar correla-

7. Tomo el ejemplo de la introducción al libro de A. Burkov, *The hundred-page machine learning book,* 2019.

ciones dentro de la información solicitada. A través de las técnicas matemáticas propias de la IA («algoritmos») se puede concluir cuál era, *en el pasado*, la respuesta correcta a la pregunta formulada. Por ejemplo, si se trata de analizar pruebas diagnósticas (ecografías) intentando detectar señales de un futuro tumor, se puede llegar a la conclusión de que las pruebas de las personas que finalmente desarrollaron la enfermedad tenían una serie de características que no estaban (o estaban mucho menos) en el resto de las personas. Una vez conseguido ese «retrato robot», que habría funcionado en el pasado, se aplica al futuro, en el sentido de que, a partir de los datos que tenemos (pruebas diagnósticas de pacientes actuales, que no sabemos si desarrollarán o no la enfermedad), el sistema predice el riesgo de desarrollo de la enfermedad en cada persona.

Los sistemas de IA, en definitiva, ayudan a tomar decisiones sobre la base del análisis matemático de una enorme cantidad de datos relativos a la experiencia anterior de esas mismas cuestiones sobre las que se debe decidir.

De aquí se deduce que, mientras que la robotización o automatización se utiliza allí donde las decisiones individuales están totalmente determinadas por el marco normativo (marco que se traduce o vuelca en un programa informático, facilitando su aplicación a los casos concretos), la IA se utiliza allí donde sí existe un margen para la decisión, puesto que la IA sirve precisamente para facilitar herramientas que orientan la decisión en un sentido o en otro. La diferencia entre un caso y otro es transparente.

El efecto práctico de la IA es establecer un «baremo» en todos esos casos, un baremo que establece una predicción (que el operador tendrá en cuenta o no) a partir de una serie de datos que la experiencia pasada ha demostrado que tienen una correlación con la cuestión que se intenta resolver. Los sistemas de IA intentan resolver una pregunta (cuya respuesta es desconocida) a partir de datos que son conocidos, sabiendo que, en el pasado, se demostró la existencia de una correlación entre los datos que conocemos y la pregunta cuya respuesta no conocemos.

Para ser conscientes del efecto práctico de la IA, hay que tener en cuenta que, antes de que la misma comenzase a emplearse, muchas de esas decisiones se tomaban de forma intuitiva por operadores humanos, a partir de su «ojo crítico» o, en general, de criterios de experiencia imposibles de objetivar. La IA no sustituye a criterios legales de toma de decisiones (no podría hacerlo), sino que opera allí donde existe un margen de decisión entregado a los operadores humanos. Esto sucede tanto en decisiones privadas (decisiones de inversión, de selección de objetivos comerciales, etc.) como públicas (por ejemplo, decisiones en cuanto a dónde se ubican los recursos que la Administración tiene a su disposición, cuáles de las múltiples empresas sujetas a control son inspeccionadas, etc.).

En otros casos, la IA permite actuar de manera diferenciada en contextos en los que con anterioridad se operaba de forma indiscriminada, tratando por igual a todos los sujetos. La IA permite, por ejemplo, que las empresas ofrezcan condiciones contractuales a medida o se dirijan preferentemente a determinados clientes potenciales, mientras que anteriormente no podían discriminar y se veían obligadas a dirigirse al público en general, perdiendo esfuerzos en muchos casos (por dirigirse también a personas que no tenían ningún interés en comprar el producto).

Un campo que ha ganado mucho protagonismo en los últimos tiempos, y que no se separa demasiado de este esquema, es el de la IA «generativa», que permite la creación de textos, audios o incluso videos que respondan a las preguntas o instrucciones del usuario. En su versión más básica, tenemos modelos similares a los correctores o predictores de textos que todos conocemos a través de los teléfonos móviles y ordenadores. A partir de un corpus de palabras y de la propia conducta anterior del usuario, el sistema «predice» cuál es la palabra que éste ha empezado a escribir y se la propone. Versiones más sofisticadas son las que proporcionan las imágenes o textos que —hasta donde llegan las capacidades del sistema— mejor se adaptan al conjunto de palabras suministrado por el usuario. Todos estos sistemas se basan en análisis masivo de información disponible en internet (vídeos, textos, audios, etc.) y en la identificación de correlaciones, que después permiten ofrecer o incluso generar la imagen o texto que mejor se ajuste a la petición.

Este tipo de sistemas pueden servir para realizar tareas tediosas o simples, que, aunque no son completamente automáticas, no tienen excesiva dificultad partiendo de unas instrucciones básicas. Son el tipo de tareas que no requieren una gran formación y que, en la práctica, tienden a recibir una baja remuneración porque no es difícil encontrar a alguien que las haga (tareas que a veces se conocen como «commodities»). Los sistemas de IA generativa pueden sustituir a operadores humanos en la realización de estas tareas.

Una segunda característica de estos sistemas es su carácter transversal o auxiliar, puesto que no están dirigidos a realizar una tarea concreta, sino que se pueden utilizar como apoyo para muchas. Esto puede dificultar el control, porque es fácil entender que se someta a unos requisitos específicos un sistema dirigido a identificar posibles infractores, y ligado a una concreta aplicación, mientras que es menos claro cuando se trata de un programa que tiene múltiples utilidades (como corregir textos de cualquier tipo).

3. Ventajas y aportaciones de la IA

La principal ventaja de la IA es la eficiencia en todas sus dimensiones, en la medida en que, en lugar de actuar a ciegas o de forma indiscriminada, quien utiliza un sistema de IA puede tomar decisiones orientadas y maximizar los resultados. El caso de la publicidad es revelador: si el mismo esfuerzo publicitario se utiliza, no para que todos los consumidores vean un anuncio, sino para que lo vean aquellos que tienen más probabilidades de comprar el producto, se podrá influir con más intensidad en esos consumidores y se obtendrán mejores resultados que con el mismo esfuerzo distribuido entre todos los consumidores, gran parte de los cuales no tiene ninguna probabilidad de comprar el producto a pesar de la publicidad. Si una prueba de diagnóstico cara e invasiva se utiliza sólo en aquellos pacientes que un sistema de IA ha identificado como probables desarrolladores de la misma, se reducirán los «negativos» y el desperdicio de recursos, se podrá estudiar mejor a los pacientes de riesgo y se reducirá también el número de enfermos no diagnosticados.

La eficiencia sirve a las empresas para obtener mejores rendimientos y también a las Administraciones Públicas para prestar mejores servicios. También permite obtener mayores avances científicos, al hacer un filtrado previo que permita identificar las hipótesis con más

probabilidades de ser ciertas, lo que a su vez permite que los experimentos (que son caros y llevan tiempo, por lo que están limitados) se concentren en esas hipótesis, en lugar de realizarse indiscriminadamente o en función de intuiciones.

Visto desde otro punto de vista, la IA permite dejar en manos de ordenadores la realización de tareas tediosas y poco creativas. Las capacidades de los sistemas de IA (que ya hemos visto), es decir, la posibilidad de hacer predicciones en entornos definidos a través de grandes cantidades de datos, permiten enseñarles a realizar tareas que, aunque no son puramente automatizables, porque en ellas pueden producirse circunstancias más o menos imprevistas, sí pueden reducirse a unas rutinas. Son tareas rutinarias, que también se dan en trabajos intelectuales, y que pueden dejarse en manos de ordenadores gracias a la IA, lo que libera a los trabajadores de una organización (o a los profesionales autónomos) para que puedan dedicarse a otras más creativas.

Por último, también destacaría en esta enumeración no exhaustiva el hecho de que la IA permite aprovechar el potencial de los datos, es decir, de la información, y supone, en cierto modo, un antídoto a las decisiones subjetivas y basadas en prejuicios o en ocurrencias. Los trabajos ya famosos de Kahneman y Twersky han explicado cómo el cerebro tiende a decidir sobre la base de respuestas rápidas y no racionales, debido a que el razonamiento es un proceso más largo[8]. Y han explicado (y me parece muy interesante) que lo que denominados «ojo clínico» o seguridad en el juicio no son más, muchas veces, que juicios sumarísimos de este tipo, con frecuencia arbitrarios. Operar con datos es algo mucho mejor y la IA puede servir para hacerlo, teniendo en cuenta que no necesariamente sustituye al juicio humano, sino que puede reforzarlo al suministrarle información y puntos de vista.

4. Riesgos e inconvenientes

La IA no está exenta de inconvenientes, en los que se está insistiendo mucho en los últimos años, hasta el extremo de iniciativas tan llamativas como la petición de una moratoria de seis meses para poder implantar alguna forma de regulación, lanzada a finales de marzo de 2023 y pronto criticada.

La IA es, como estamos viendo, una técnica predictiva y de toma de decisiones. El primer riesgo es que las predicciones estén equivocadas. Esto es algo común a muchas otras técnicas que se han presentado como capaces de predecir algo, desde el tiempo atmosférico hasta las cotizaciones bursátiles. Se trata de un riesgo *para la técnica predictiva en cuestión* (que sería abandonada al constatarse sus fallos), no tanto para la humanidad, salvo que ésta se coloque voluntariamente en manos de esa técnica predictiva. Por tanto, el riesgo estaría en una utilización de la IA indiscriminada y carente de cautelas, no tanto en la propia IA.

Son muchos los factores concretos que pueden provocar que los sistemas de IA emitan predicciones equivocadas. En primer lugar, los datos de partida pueden ser erróneos o in-

8. En particular, Kahneman, D., *Thinking, fast and slow,* Penguin, 2012 (el libro se publicó originalmente en 2011; hay traducción española: *Pensar rápido, pensar despacio,* Debate, 2012).

suficientes y eso puede lastrar el resultado, como es fácil comprender. También puede suceder que un sistema creado a partir de datos de una determinada realidad (un país, una comunidad autónoma) se utilice en otro contexto diferente. Es fácil entender que pueden producirse errores porque el sistema no conoce circunstancias diferenciales que pueden afectar al resultado. Por otro lado, los sistemas de IA no surgen de forma automática a partir de los datos, sino que son creados por un equipo humano que tiene que tomar muchas decisiones al diseñarlo: seleccionar qué datos son relevantes para responder al problema formulado, «limpiarlos», combinarlos de forma adecuada para compensar posibles insuficiencias y carencias de la muestra, etc. En cierto modo, es como la labor del enólogo, que no produce vino de forma mecánica a partir de la uva, sino que puede producir vinos muy diferentes con una misma uva. El sistema de IA siempre «predice» acertadamente el pasado, en el sentido de que, dados los datos del pasado, su respuesta es coherente con lo que sucedió en el pasado, pero todas esas circunstancias pueden hacer que su predicción no sea acertada para el futuro[9].

El siguiente riesgo es el que normalmente se aborda diciendo que la IA puede perjudicar a las personas más vulnerables. Recordemos que la IA permite que algunas cosas que se hacían indiscriminadamente pasen a realizarse de una forma selectiva, identificando unos objetivos en los que tiene más sentido actuar. En principio, esta selección se realiza sobre la base de criterios objetivos (inspeccionar a las empresas que todo indica que es más probable que estén cometiendo una infracción), pero se puede producir el resultado de que determinadas personas o colectivos resulten injustamente seleccionados y sometidos a una presión mayor.

Un problema que exige un amplio desarrollo es el riesgo de que el sistema de IA detecte una correlación entre aquello que se está buscando y un factor como la raza o la religión. Pensemos en un sistema de selección de empleados (utilizado para filtrar las solicitudes y determinar el paso a la entrevista) que detecte que los empleados de una determinada religión han tenido un peor comportamiento en el pasado. Se daría una puntuación negativa a los aspirantes de esa religión. Parece claro que ese resultado no cumpliría el test del artículo 14 CE (independientemente de que seguramente se podría concluir, en una verificación, que ese juicio es, además, incorrecto porque se basa en una muestra equivocada o porque no tiene en cuenta circunstancias concomitantes que son las que explican ese resultado, y no la religión). El sistema debería ser, por tanto, «ciego» ante ciertos factores para evitar convertirse en discriminatorio.

El problema no se resuelve tan fácilmente, porque, a veces, se elimina un factor (religión, raza) y vuelven a detectarse correlaciones que indirectamente llevan al mismo resultado. Por ejemplo, quitamos los datos de religión pero el sistema detecta una correlación entre los aspirantes procedentes de un barrio y una característica negativa, y empieza a puntuar de forma desfavorable a los procedentes de ese barrio, que coincide que es el habitado ma-

9. Por ejemplo, un sistema de predicción del riesgo de insolvencia basado en datos del pasado producirá un baremo para valorar a los potenciales clientes a partir de sus datos, que permitirá clasificarlos en función de su nivel de riesgo. El sistema «funciona» en el sentido de que, dados los datos de clientes pasados, el nivel de solvencia asignado a cada uno encaja con su conducta real (es decir, se clasifican como poco solventes a los que finalmente no devolvieron el préstamo). Eso no significa que vaya a funcionar en el futuro.

yoritariamente por personas de esa religión. Nos encontramos el mismo problema. Al igual que sucede con el denominado «olvido oncológico» y la discriminación de los portadores del VIH, nos encontramos ante una línea fina entre la discriminación y la existencia de circunstancias objetivas que pueden justificar un tratamiento diferente[10]. Creo que al final es necesario plantearse qué decisiones es necesario tomar de forma indiscriminada y en cuáles se pueden implantar criterios de preferencia, y también ver qué grado de justificación se exige para tales criterios.

Además, cuando se toman decisiones de forma dirigida o polarizada gracias a sistemas de IA (por oposición a cuando esas decisiones se toman de forma indiscriminada), se puede producir un efecto distorsionador, según la frase evangélica «quien busca, halla». Parece que la realidad confirma la predicción cuando tal vez también se habría hallado lo que se busca (por ejemplo, una infracción) en otro lugar si se hubiera buscado en él («profecía autocumplida»). Puede convertirse un indicio en una realidad y hacer que, por ejemplo, determinados colectivos sean considerados peores solventes o más infractores sólo porque han sido objeto de una pesquisa más cuidadosa que ha encontrado infracciones y que no se ha llevado a cabo en otros lugares.

Los sistemas de IA no sólo pueden cometer errores, sino que esos errores pueden ir en una determinada dirección y perjudicar a determinadas personas o colectivos. Los que tienen menos capacidad de defensa tienen más probabilidades de ser perjudicados en general y también por la IA. Y también puede suceder que los estereotipos socialmente difundidos, que también afectan o pueden afectar a los creadores de sistemas de IA, se introduzcan en los sistemas a través de las decisiones tomadas en su diseño.

Todos estos riesgos no deben hacernos olvidar que las decisiones tomadas por operadores humanos sin interferencia de la IA también están llenas de sesgos y muchas veces sus razones verdaderas no son las que aparecen en la motivación. Se podría motivar jurídicamente una decisión y también la contraria, y las razones de que la decisión sea una u otra son ocultas. Los sistemas de IA pueden cometer errores o estar mal diseñados, pero, por oscuros que sean, son mucho más transparentes que el comportamiento humano.

Otro de los riesgos asociados a la IA (que, como suele suceder, es la otra cara de una de sus ventajas) es la destrucción de puestos de trabajo, puesto que puede o podría realizar tareas que actualmente realizan humanos. Este argumento «ludita» es el que me parece menos relevante. Hay sectores, como la administración de justicia o grandes ramas de las administraciones públicas, que se encuentran crónicamente infradotadas de personal pese a lo que

10. *Vid.* Muñoz Paredes, Mª. L., *El deber de declaración del riesgo en el seguro,* Aranzadi, Cizur Menor, 2018, págs. 65-68; «La discriminación de los asegurados en el precio del contrato fijado con uso del big data», en Girgado Perandones, P./Gonzáles Bustos, J. P. (coor.), *Transparencia y competitividad en el mercado asegurador: Insurtech, distribución, protección del cliente, seguro marítimo y pandemia,* Comares, Granada, 2021, págs. 263-296, especialmente 278-279. Como recuerda la autora, la Ley 4/2018, de 11 de junio, de reforma simultánea de la Ley de Consumidores y Usuarios (LCU, Texto Refundido aprobado por Real Decreto Legislativo 1/2007) y de la LCS, aprobada para luchar contra esta discriminación, prohíbe a los aseguradores imponer «condiciones más onerosas, por razón de tener VIH/SIDA u otras condiciones de salud, *salvo que se encuentren fundadas en causas justificadas, proporcionadas y razonables, que se hallen documentadas previa y objetivamente»,* admitiendo, por tanto, un trato diferente basado en circunstancias de esta naturaleza. Actualmente hay que tener en cuenta el Real Decreto-ley 5/2023, de 28 de junio, que modifica el artículo 10 y la Disposición Adicional 5ª de la Ley del Contrato de Seguro. La Disposición Adicional 5ª mantiene el texto entrecomillado.

pueda parecer. Sólo así se explican las grandes demoras en la justicia o, por ejemplo, en el reconocimiento del ingreso mínimo vital. Por lo tanto, que los actuales empleados puedan tener una ayuda que les quite trabajo tedioso es una ventaja. Otra cosa es la necesidad de gestionar adecuadamente este proceso, para evitar que, como ha sucedido en gran medida con la digitalización, se traduzca en más trabajo para los técnicos, que en muchos casos se han visto obligados a asumir tareas administrativas supuestamente robotizadas (introducir datos en una aplicación informática), sin que ello haya beneficiado aparentemente a nadie, puesto que no ha ahorrado personal administrativo ni ha permitido que éste se dedique a otras tareas.

En todo caso, es una evidencia empírica que ningún avance tecnológico se ha detenido por la destrucción de puestos de trabajo: se mecanizó la agricultura (expulsando a miles de trabajadores a las ciudades), se mecanizó la industria (con parecidos resultados) y es lógico que se mecanicen los servicios, incluidos los intelectuales, que es lo que en cierto modo supone la IA. Si llegáramos al escenario extremo de que un alto número de personas no pudieran encontrar un empleo productivo, habría que arbitrar alguna solución para que también tengan ingresos y se beneficien del trabajo realizado por la IA, algo que no es totalmente nuevo en la historia, en la que hay supuestos de sociedades cuyos ciudadanos apenas trabajaban, bien por beneficiarse del trabajo esclavo (Roma, en algunos momentos), bien porque la mayor parte del trabajo lo realizan inmigrantes (como ocurre en algunos países árabes productores de petróleo).

También se alude con frecuencia al peligro de que los productos de la IA se vuelvan indetectables, en dos direcciones: imágenes o audios que parecen reales pero no lo son (de modo que parece que ha sucedido algo que es falso o que alguien ha dicho algo que no ha dicho, con posibles repercusiones políticas o incluso penales) y trabajos que no ha realizado quien los presenta como suyos, sino que son el producto de un sistema de IA (típicamente, estudiantes que presentan como propios trabajos de este origen).

En cierto modo, esa capacidad de «engaño» es una prueba de la validez de la IA. Cualquier herramienta, para funcionar bien, ha de superar a los humanos en su campo de acción. Sería más cuestionable o peligroso si la IA suplantase a los humanos, no en tareas auxiliares, sino en las más importantes, porque en ese caso éstos se verían desplazados e inútiles. Sin embargo, no parece que realizar trabajos consistentes en un mero acarreo de internet sea lo que nos define como humanos, aunque, lógicamente, no se pueden poner puertas al campo ni saber qué hará en el futuro la IA. Por otro lado, el hecho de que la IA pueda crear imágenes o audios que parezcan reales no siéndolo en realidad nos obliga a establecer más cautelas a la hora de dar por supuesta la veracidad de una imagen o un audio (como ocurrió hace años con los programas de manipulación de fotos e imágenes) y a dejar de suponer, en definitiva, que una imagen fotográfica (al contrario que un dibujo o una pintura) es real.

Otro de los riesgos muy frecuentemente mencionados es el de que la IA, al ofrecer a cada usuario de redes sociales los mensajes o noticias que son más acordes con su manera de pensar (expresada en su actividad anterior en internet: páginas visitadas, reacciones en redes sociales, etc.), lleva a la formación de «burbujas» o «campanas de eco», de modo

que cada persona percibe una realidad distorsionada, en la que sólo aparecen noticias o mensajes que reafirman sus ideas o prejuicios, lo que contribuiría a la polarización y la desinformación. También aquí encontramos un punto de exageración. En la sociedad «tradicional» en la que las personas se informaban a través de periódicos, radios o cadenas de televisión, también era frecuente que cada persona se informara a través de los medios más acordes con sus preferencias y que tuviera esa misma versión distorsionada de la realidad. De hecho, internet facilita una mayor «promiscuidad» informativa, en la medida en que es más fácil tener acceso, aunque sea superficial, a webs o noticias de distintos medios de comunicación. En todo caso, lo que se ha producido es una pérdida de poder de los medios de comunicación tradicionales, que han perdido el monopolio en la intermediación informativa, pero esos medios también pueden manipular o distorsionar la realidad.

Con frecuencia se agita el relato de una IA que logra superar a los humanos hasta casi tomar el control. Precisamente, la petición de una moratoria de seis meses en el desarrollo de la IA, que surgió tras el despegue deslumbrador de Chat-GPT, se basaba en esa idea de que una IA que pudiera manejar el lenguaje dominaría a los humanos y, por ejemplo, en poco tiempo determinaría sin resistencias el desenlace de las elecciones (pensando, por supuesto, en las elecciones presidenciales de USA).

Creo que aquí se produce un desenfoque bastante habitual, el de ver a la IA como un ente personificado cuando en realidad es un instrumento que algunos humanos utilizan para ejercer su poder (especialmente económico, pero poco también en otros campos) sobre otros humanos. La IA es un instrumento que permite a quien tiene el poder ejercerlo de una forma todavía más intensa y eficaz, perfilándolo y dando un trato diferente a los sometidos a ese poder, lo que le permite, por un lado, conseguir mayores resultados (la semilla no se lanza al vacío, sino que se dirige a los lugares más fértiles) y, por otro, tratar a cada persona en función de sus datos y su historial de navegación, lo que supone un gran ataque a la intimidad. Pero «la IA» no es un ser artificial que toma el control, sino un instrumento del que se sirven determinados humanos.

En fin, el último riesgo que quiero mencionar es que los humanos pasan a encontrarse en una situación nueva, en la que cada acción u omisión que realizan en un entorno digital (y son prácticamente todas) puede ser utilizada en su contra, es decir, quede registrada y pueda ser utilizada por empresas para clasificar a esa persona y darle un trato u otro en función de ese perfil. Cuando digo «dar un trato u otro» me refiero a la información que se le ofrece en primer lugar, a las condiciones contractuales, a la publicidad que recibe, etc. Cuando las empresas presionan a los usuarios para que «descarguen la aplicación» y se registren, o para que utilicen tarjetas de fidelización, uno de los objetivos es tener mucho más conocimiento de esa persona y controlar mejor el tratamiento que se le da.

Por supuesto que existen límites para ese perfilado, aunque su funcionamiento es imperfecto. Los controles establecidos por la legislación de protección de datos pueden ser levantados por el consentimiento del interesado, que se le exige de manera rutinaria al entrar en cualquier página web, por lo que es frecuente que ese consentimiento se preste también de forma rutinaria. El proyecto de Reglamento de IA prohíbe el establecimiento de sistemas

de recompensa, pero eso no impide completamente el perfilado, que es una técnica muy extendida en la IA.

En definitiva, el riesgo está en el denominado «capitalismo de control», en el que internet deja de ser un escaparate utilizado por cualquier operador para acceder a un público, y se convierte también en un mecanismo por el que los «espectadores» entregan información relevante sobre sí mismo que puede ser explotada por quienes ofrecen cosas, obteniendo así mejores resultados[11].

5. Encaje jurídico

A partir de aquí, se trata, no ya de perfilar qué regulación debería darse (en su caso) a la IA, sino de contribuir a establecer cómo ha de ser tratada jurídicamente.

El punto de partida, esbozado ya en el apartado anterior, es que la IA es un instrumento para la toma de decisiones (o, por decirlo de otro modo, para el ejercicio del poder). No actúa, por supuesto, por su cuenta, al tratarse de sistemas creados por unos proveedores y empleados por unos usuarios con la finalidad de tomar mejores decisiones y obtener mejores resultados (normalmente en relación con unos «sujetos» finales, a los que se explota —o se atiende— de forma más eficiente). Quienes ejercen el poder no pueden esconderse detrás de la IA porque ésta no ejerce ningún poder, sino que es un instrumento que les ayuda a hacerlo.

Aunque hasta ahora se ha incidido sobre todo en regular la IA para que sus predicciones sean acertadas (por ejemplo, asegurando la calidad de los datos) y para evitar que produzcan efectos discriminatorios, lo cierto es que es necesario abordar esta vertiente de la IA como instrumento para el ejercicio de poderes.

En un Estado de Derecho, los poderes (públicos y privados, derivados de leyes o derivados de contratos) tienen unos límites, cuyo cumplimiento vigila el poder judicial o la Administración. Así, hay límites negativos o externos (el acreedor no puede pedir una cantidad superior a la que se le adeuda) y límites internos (para cambiar la calificación urbanística es necesario aportar una justificación exhaustiva). Cuando se emplea IA como instrumento para el ejercicio del poder, estos requisitos se aplican de la misma manera.

Así, cuando un empleador utiliza IA para determinar la retribución de los trabajadores (dentro de los límites en que el convenio o el contrato dejan en manos del empresario la posibilidad de concretar las retribuciones o distribuir los tiempos de trabajo), se aplica el mismo marco que cuando esa decisión se toma sin el apoyo de un sistema de IA. Por tanto, lo primero que hay que preguntarse es si las concretas opciones tomadas a partir del «informe» de la IA son, o no, admisibles dentro del marco jurídico de ese poder empresarial. Y, en segundo lugar, si están adecuadamente justificadas.

Para saber si la IA puede utilizarse para tomar una determinada decisión, hay que ver, repito, cuál es su marco jurídico, incluida la forma de justificarla. Las decisiones que afectan

11. Expresión debida a S. Zuboff, *The age of surveillance capitalism,* Profile Books, London, 2019.

a terceros están siempre sometidas a control (en último término, un control judicial), en el que se verificará si esas decisiones exceden el marco y si están adecuadamente justificadas o motivadas de acuerdo con los criterios con que, según ese marco normativo, deben fundamentarse. Así, por ejemplo, el cese de un funcionario será legal si ocupa un puesto de libre designación (si el puesto se ocupa en virtud de concurso, no cabe el cese discrecional) y si se encuentra debidamente motivado.

¿Puede servir la IA como motivación? De nuevo debemos acudir al marco jurídico. La mayoría de las decisiones sólo pueden adoptarse cuando se constata la producción de un supuesto de hecho fáctico. En estos casos, la IA no nos sirve, porque ésta constata una probabilidad, nos ofrece una *predicción*, pero no se puede confundir con la verificación de un presupuesto de hecho.

La IA puede ser más útil, paradójicamente, para justificar decisiones cuyo presupuesto de hecho (es decir, lo que debe constatarse para que la decisión sea legítima) es una situación de *riesgo*, una *probabilidad*, porque son esa clase estados los que puede predecir o constatar de forma solvente un sistema de IA. Pensemos, por ejemplo, en la determinación de las zonas sometidas a un riesgo de inundación, que es relevante en la planificación urbanística, o en el riesgo de que una fusión empresarial provoque una reducción de la competencia, que también es un hecho relevante a efectos de autorizar, o no, la fusión. Son cosas que sí pueden acreditarse con la ayuda de un sistema de IA.

Las normas sectoriales pueden referirse a la IA. Normalmente, no se regula ni se limita el tipo de técnicas con que se puede probar el hecho determinante de la decisión (principio de libertad de prueba, válido tanto en el orden judicial como en el administrativo). La IA podrá, por tanto, emplearse, sometida, como cualquier otro medio de prueba, a verificación contradictoria, lo que exigirá informar de manera suficiente sobre el funcionamiento del sistema de IA, para que pueda verificarse el acierto de sus predicciones (sin esperar a que se cumplan, claro).

El hecho de que la mayoría de las decisiones jurídicas (administrativas, judiciales) tengan que basarse en la constatación de hechos (y no en predicciones) hace que los sistemas de IA desplieguen su mayor utilidad, no para la toma decisiones en sentido estricto (actos administrativos, resoluciones, sentencias) sino para actos preparatorios, tanto en el contexto de la actividad jurídica de la Administración (la que concluye con actos administrativos) como en el de la actividad material (actividad asistencial, prestación de servicios públicos en sentido amplio como sanidad, educación, etc.). Me refiero a decisiones como: distribuir los recursos de que dispone la Administración (dónde situar más personal o medios materiales y predecir así los picos de demanda), señalar objetivos para tareas como la inspección de posibles infracciones, indicar qué personas pueden encontrarse en una situación de riesgo que determine que se les ofrezca un determinado servicio o prueba diagnóstica, etc. En el ámbito privado (decisiones de autoorganización empresarial o adopción de políticas comerciales), como no se afecta a derechos de terceros (nadie tiene derecho a que se le dirija publicidad o a que se abra una tienda en su ciudad), las exigencias jurídicas son menores. La IA tiene, como estoy diciendo, esa función *aproximativa* (pasar de 100 a 10, por decirlo gráficamente) más que resolutoria. Lo normal es que la IA evidencie *indicios* y que éstos

den lugar a una investigación administrativa de la que se deducirán, en su caso, las *pruebas* que son necesarias para que se tome la decisión.

Muchas de estas decisiones aproximativas están sometidas a un marco jurídico poco estricto, en el que no se exige apenas motivación. Son actos administrativos de trámite, no recurribles autónomamente, y, si se constata que la resolución del procedimiento es materialmente correcta (porque concurre su presupuesto de hecho), es difícil que un problema en el acto de iniciación dé lugar a la anulación de la resolución.

En mi opinión, el hecho de que en esa decisión sometida a un marco jurídico poco estricto se haya utilizado IA no debería llevar a que se le impongan exigencias jurídicas considerablemente superiores.

6. El proyecto de Reglamento de IA: una síntesis telegráfica

Tras una profusión de declaraciones no vinculantes y de recomendaciones «éticas», el proyecto de Reglamento de la UE sobre IA (denominado «Ley de IA» o *AI Act*) es el texto normativo más sólido y ambicioso que existe en el panorama comparado, por lo que es lógico que concentre todas las miradas. Su análisis exigiría un espacio que no está disponible en este momento, por lo que me concentraré en algunas pinceladas.

El Reglamento parte de una definición de la IA que se corresponde con la que se está utilizando en este trabajo. La versión publicada por el Consejo en diciembre de 2022 lo define como «un sistema concebido para funcionar con elementos de autonomía que, a partir de datos e información generados por máquinas o por seres humanos, infiere la manera de alcanzar una serie de objetivos, utilizando para ello estrategias de aprendizaje automático o estrategias basadas en la lógica y el conocimiento, y produce información de salida generada por el sistema, como contenidos (sistemas de inteligencia artificial generativa), predicciones, recomendaciones o decisiones, que influyen en los entornos con los que interactúa el sistema de IA». En la versión inicial de la Comisión (2021) no sólo no se incluía ninguna referencia a los sistemas de inteligencia artificial generativa, sino que se trataba de concretar el concepto mediante la referencia a una serie concreta de técnicas matemáticas incluidas en un Anexo (que podía actualizar la Comisión). En todo caso, está claro que se necesitan «elementos de autonomía» (queda fuera la simple automatización), que se basa en datos, que utiliza técnicas matemáticas para su manejo y que produce «información de salida» como predicciones.

El enfoque del Reglamento es tratar a la IA como una tecnología de riesgo, concretamente para «la salud, la seguridad o los derechos fundamentales», según la fórmula utilizada constantemente en la norma. La idea es que, ante la existencia de este riesgo, no resulta suficiente un enfoque **represivo** (es decir, la responsabilidad penal o civil de quien a través del sistema de IA causa daños a esos bienes jurídicos, que debe actuar como elemento disuasorio), sino que es necesario un enfoque **preventivo**, que consiste en exigir la adopción de medidas que minimicen las probabilidades de que se produzcan esos daños. Puede decirse

que, mientras que los daños corporales o materiales que pueden producir otras tecnologías del riesgo (automóviles, centrales nucleares, maquinaria en general) son claros y visibles, aquí esos daños no están tan perfilados (si dejamos a un lado el supuesto de la discriminación, que es el más claro y el que ha centrado hasta ahora la mayor parte de las construcciones doctrinales).

El Reglamento enumera prácticas prohibidas, aunque algunas de ellas son supuestos prácticamente marginales. Uno de ellos es la IA subliminal que puede llevar a que las personas se causen daños a sí mismas. También otros sistemas de IA que, aunque no sean subliminales, aprovechan las vulnerabilidades individuales o de grupo para que, de nuevo, las personas se causen daño a sí mismas.

También se prohíben los sistemas de «puntuación» o clasificación social, es decir, los que sirven para «de evaluar o clasificar a las personas físicas durante un período determinado de tiempo atendiendo a su comportamiento social o a características personales o de su personalidad conocidas o predichas, de forma que la puntuación ciudadana resultante provoque una o varias de las situaciones siguientes», a saber: que la puntuación así obtenida determine «un trato perjudicial o desfavorable hacia determinadas personas físicas o grupos de personas físicas en contextos sociales que no guarden relación con los contextos donde se generaron o recabaron los datos originalmente», o que provoque «un trato perjudicial o desfavorable hacia determinadas personas físicas o grupos de personas físicas que es injustificado o desproporcionado con respecto a su comportamiento social o la gravedad de este». Habrá que seguir la interpretación de este concepto, puesto que ya sabemos que el perfilado es una de las principales utilidades o funciones de los sistemas de IA. Si la prohibición se limita a sistemas *quasi* públicos que establezcan una puntuación general para los ciudadanos, tendrá escasa aplicación. Si se realiza una interpretación amplia, podría tener una aplicación mucho más grande (aunque no parezca probable).

El último de los supuestos prohibidos, que es el de los sistemas de identificación biométrica en tiempo real, más que una prohibición es un establecimiento de requisitos, puesto que se permite su uso si es para la prevención de delitos y con arreglo a una serie de condiciones y garantías.

Dejando a un lado esos supuestos prohibidos, se crea una categoría denominada «sistemas de IA de alto riesgo», que recoge los que se identifican en un anexo (actualizable por vía reglamentaria, sin modificar el Reglamento). Son sistemas de IA que, o bien se utilizan en conexión con actividades o productos sujetas a regulación precisamente por su riesgo (por ejemplo, sistemas de IA que sirven para manejar infraestructuras críticas, por ejemplo), o bien se utilizan con finalidades o funciones especialmente sensibles (como la justicia, el empleo, el acceso a la educación, la sanidad, etc.). Implícitamente, el Reglamento está admitiendo que la IA se utilice prácticamente para todo, puesto que en esos casos se tratará de sistemas de alto riesgo, sujetos a unos requisitos más rigurosos, pero no por ello queda prohibido su uso.

A los sistemas de IA de alto riesgo se les ponen distintos requisitos, relativos a la calidad de los datos, a la transparencia, a la información que hay que suministrar a sus usuarios, o a la necesidad de garantizar el control humano. Se debe garantizar, en la medida de lo posi-

ble, la solidez del sistema y su ciberseguridad. En particular, es necesario establecer un sistema de gestión de riesgos, en el que se examinen desde el principio cuáles son las posibles desviaciones, errores o, en definitiva, daños que pueden producirse y se adopten medidas preventivas adecuadas. También es necesario que quede constancia (registro) del funcionamiento del sistema, a modo de «caja negra» (como las de los aviones) que permita comprobar si se han producido errores o desviaciones.

El Reglamento no establece unas normas fijas, unas medidas obligatorias que deban tomarse en todo caso, sino que exige que se adopten las medidas adecuadas y proporcionadas para minimizar los riesgos, sabiendo que no se exige que el riesgo sea inexistente (eso exigiría probablemente renunciar a la IA) y que no en todo caso será necesario o conveniente adoptar las mismas medidas, puesto que éstas serán proporcionales al riesgo generado en cada caso y también a otros factores como el tamaño de la empresa. Es una forma de regulación seguramente inevitable pero distinta de la más habitual en sectores cuya regulación es más antigua. Por poner un ejemplo, es como si la normativa de circulación no estableciera una velocidad máxima concreta, sino que simplemente dijera que se deberá adaptar la velocidad a las circunstancias del tráfico.

La forma de garantizar el cumplimiento de estas exigencias «principales» también merece un comentario. No se trata, desde luego, de un control público en forma de autorización. En algunos casos, el sistema de alto riesgo debe ser verificado por una entidad certificada, es decir, una entidad (que puede ser privada) que interviene a petición del proveedor, es decir, el creador del sistema de IA. Se trata de una técnica que se aplica en muchos otros campos (la ITV, por ejemplo). Aunque el certificador es retribuido por el proveedor (es decir, el controlado paga al controlador), se establecen medidas dirigidas a evitar conflictos de intereses.

En otros casos opera la autovigilancia, el *compliance*. El proveedor debe cumplir esas reglas y dejar constancia razonada de cómo lo hace, para que la Administración competente pueda, en su caso, verificar ese cumplimiento. El mecanismo de garantía final es la posible imposición de sanciones administrativas. En los casos en que no se cumplan las obligaciones formales (es decir, que el proveedor no haya realizado la gestión de riesgos, o no haya sometido el sistema a certificación externa si está obligado a ello), es fácil justificar la imposición de sanciones. Si el problema es la interpretación de las garantías exigibles, es decir, si el proveedor ha traducido correctamente los principios generales (seguridad, solidez, transparencia, etc.) que impone el Reglamento, será más difícil hacer compatible la sanción con el principio de tipicidad, visto que la definición concreta de la infracción se hace sólo en la resolución que impone la sanción.