

Resumen

La importancia creciente de la inteligencia artificial (IA) y la robótica en nuestra sociedad es ya una realidad indiscutible. Sin embargo, a pesar de los importantes avances obtenidos en la última década, el despliegue de estas tecnologías es ciertamente asimétrico y su omnipresencia en nuestras vidas genera controversia. En este artículo se intentan analizar las claves de la situación actual de estas tecnologías bajo el prisma de la innovación y en perspectiva europea, con un acento final español. El creciente poder de los grandes gigantes tecnológicos y el aparente retraso que Europa tiene respecto a ellos incita a la reflexión sobre las causas de ese desequilibrio y, sobre todo, acerca de una posible estrategia que mejore la competitividad del viejo continente.

Palabras clave: inteligencia artificial, robótica, aprendizaje profundo, gobernanza de datos, innovación tecnológica, estrategia nacional de inteligencia artificial.

Abstract

The growing importance of artificial intelligence (AI) and robotics in our society is already an indisputable reality. However, despite the significant advances made in the last decade the deployment of these technologies is certainly asymmetric and their omnipresence in our lives causes controversy. This article attempts to analyse the keys factors behind the current situation of these technologies in terms of innovation and from a European perspective, with a final Spanish accent. Indeed, the growing power of the big technological giants and Europe's apparent lag behind them prompts reflection on the causes of this imbalance and, above all, on a possible strategy to improve Europe's competitiveness.

Keywords: artificial intelligence, robotics, deep learning, data governance, technological innovation, national strategy on artificial intelligence.

JEL classification: O31.

ROBÓTICA E INTELIGENCIA ARTIFICIAL MÁS HUMANAS Y SOSTENIBLES

Jorge VILLAGRA

Centro de Automática y Robótica (CSIC-UPM)

I. INTRODUCCIÓN

EN los últimos años los términos inteligencia artificial (IA) y robótica han irrumpido con fuerza en el lenguaje de la calle y son objeto de discusión desde múltiples perspectivas. La «democratización» de estas tecnologías en el debate público está asociada al crecimiento exponencial de servicios que las usan, tanto en ámbitos profesionales como de consumo personal. Este auge se apoya, sobre todo, en tres grandes pilares: i) la disponibilidad de una cantidad ingente de datos, inimaginable hace solo un par de décadas; ii) la irrupción del *aprendizaje profundo* y sus diferentes variantes; y iii) la aparición de sistemas de computación con capacidades de cálculo paralelo que explotan al máximo las dos circunstancias anteriores.

Este crecimiento de capacidades, visualizado en nuestro día a día a través de espectaculares resultados en aplicaciones tan diversas como el juego Go, el procesamiento natural del lenguaje en herramientas populares como Alexa o Siri, o la percepción y toma de decisiones asociada a la conducción autónoma, genera, sin embargo, controversia.

En efecto, existe en la actualidad una serie de asuntos que orbitan en torno a la relación del humano con la IA y la robótica que inquietan a una parte importante de las sociedades avan-

zadas. Esa preocupación, si cabe más acuciante aún en el entorno europeo por el aparente retraso de Europa en relación a los *big players* mundiales, son el punto de partida de este trabajo. En este artículo se aspira a hacer un resumen de la situación actual, a describir las tendencias que se vislumbran a nivel global y a proponer diversos ejes de actuación que: i) minimicen la percepción negativa de estas tecnologías; y ii) posibiliten un posicionamiento estratégico competitivo de Europa en este campo.

II. ¿DE QUÉ ESTAMOS HABLANDO?

Es la actualidad es frecuente oír hablar de inteligencia artificial en entornos tan diversos como los debates políticos, los anuncios de lavadoras o los programas educativos. En algunos casos se aprovecha su potencia como tecnología transformadora para reclamar la exclusividad de un producto: en otros, en cambio, es vista como una seria amenaza, no solo al *status quo*, sino a una evolución sostenible del mercado de trabajo o incluso de la posición central del humano en el mundo que vivimos.

Para poder diagnosticar el estado de la tecnología, evaluar su impacto presente y pronosticar su evolución futura, es necesario definir el alcance de la misma y esbozar una mínima ontología.

John McCarthy, comúnmente conocido como el padre de la IA, escribió en 1983: «Es difícil ser riguroso sobre si una máquina realmente “sabe”, “piensa”, etc., porque nos resulta difícil definir estas cosas. Comprendemos los procesos mentales humanos solo un poco mejor de lo que un pez comprende la acción de nadar» (McCarthy, 1983). Esta cita resume la dificultad tradicional para delimitar el ámbito de influencia de la inteligencia artificial. Aún así, unos años después, el propio McCarthy (1998) se atrevió a dar la siguiente definición: «Es la ciencia que se ocupa de diseñar y fabricar máquinas inteligentes, principalmente a través de programas informáticos. Aunque a menudo intenta emular la inteligencia humana, no tiene por qué limitarse a métodos biológicamente observables». De esta visión, que sigue siendo extraordinariamente general, se infieren dos aspectos relevantes:

- la IA no tiene porque verse únicamente en una app en un móvil o a través de la nube; puede ir embebida en un sistema físico con cierto grado de movimiento en el mundo que le rodea (un robot, de los que se hablará algo más adelante);
- la clave para entender qué diferencia, por ejemplo, una muñeca moderna, que repite sistemáticamente las mismas frases, de un sistema que juega mejor que Gary Kasparov al ajedrez está en el uso de «inteligencia». ¿Y que es inteligencia?

El propio McCarthy, aun confesando su escepticismo sobre una buena definición de inteligencia humana: «No veo que la inteligencia humana sea algo

que los humanos puedan llegar a entender», la define de manera general como la capacidad de alcanzar objetivos en el mundo. Para ello, es necesario disponer como mínimo de habilidades en lógica, comprensión del mundo, aprendizaje, razonamiento, planificación y resolución de problemas. En diferentes grados, estos rasgos los comparten muchos animales y algunos programas informáticos. Sin embargo, en la inteligencia humana emergen otras capacidades singulares que todavía están lejos de alcanzarse por las máquinas, y, por tanto, nos mantienen alejados de la singularidad tecnológica (Shanahan, 2015): la autoconciencia, el conocimiento emocional, la creatividad, la inteligencia social o el pensamiento crítico.

Desde principios de los años sesenta del pasado siglo, la comunidad científica ha ido generando conocimiento en relación al primer grupo de habilidades. La producción científica ha tenido altibajos, atravesando incluso dos épocas de desilusión tras grandes expectativas generadas por los avances en años previos (los llamados inviernos de la IA en las décadas de 1970 y 1990). Sin embargo, a mediados de los años 2010, la Universidad de Toronto presenta el primer trabajo sobre *aprendizaje profundo* (Lecun, Bengio y Hinton, 2015) y, desde entonces, los avances no han cesado, no solo en forma de una explosión de resultados y de capacidades, sino, sobre todo, en su despliegue masivo en usos comerciales y profesionales. Desde entonces, los hitos se han ido sucediendo vertiginosamente (e.g., la herramienta de generación automática de contenidos GPT3 o los avances de Waymo en conducción autónoma). Los resultados son tan espectacular

res que numerosas empresas de prácticamente todos los sectores se animan a desplegar soluciones inspiradas por esos avances («No pregunte qué está cambiando la IA, pregunte qué no está cambiando la IA», Oliver, 2018). Diversas personalidades influyentes del mundo empresarial e incluso de la política muestran un entusiasmo desbordante e incluso una visión utópica habilitada por la IA. Sin embargo, es necesario matizar que existe una especie de sinécdoque colectiva e inconsciente, puesto que la inmensa mayoría de los avances actuales y de los que llegarán en los próximos años se nutren de la explotación extensiva del *aprendizaje profundo*, que no es más que un tipo de *aprendizaje máquina*, rama de la inteligencia artificial.

III. APRENDIZAJE PROFUNDO: ESTADO ACTUAL, LIMITACIONES Y OPORTUNIDADES

Según Perrault *et al.* (2019), antes de 2012, los resultados de la IA seguían de cerca la ley de Moore, doblándose cada dos años. Después de 2012, la capacidad de cómputo de las redes neuronales (elemento base del *aprendizaje máquina*) se duplica cada 3,4 meses. Ese incremento, sustentado en los crecimientos exponenciales de datos disponibles y en el aumento de la capacidad de cómputo, ha empezado a generar resultados inimaginables hace pocos lustros, generando expectativas impredecibles para los próximos años. Tómense como ejemplos: i) el *Stanford Question Answering Dataset* (Rajpurkar, Jia y Liang, 2018), conjunto de datos relacionados con la comprensión lectora, que consiste en preguntas aleatorias

planteadas sobre un conjunto de artículos de Wikipedia; o ii) el reto VQA que pone a prueba la capacidad de los ordenadores para razonar interpretando imágenes, a través de la visión artificial y del procesado natural del lenguaje. En el primer caso, la máquina superó al humano hace ya dos años; en el segundo, las prestaciones promedio de la IA son en 2021 prácticamente iguales a las del hombre.

Estos datos espectaculares (que, incorrectamente transmitidos, pueden llegar a generar cierto desasosiego) no deben ocultar la realidad compleja sobre la que reposan. En primer lugar, como se ha esbozado anteriormente, la inteligencia se articula en torno a diferentes capacidades; el aprendizaje es una muy importante, pero el resto de ellas no ha emergido (ni se prevé que lo haga en un corto plazo) únicamente a partir de la acumulación masiva de datos y de la mejora de las capacidades de cómputo. En efecto, las muy buenas prestaciones que se consiguen hoy en diferentes *benchmarks* son, casi de manera general, respuestas optimizadas a problemas de reconocimiento de patrones para los que cada algoritmo ha sido entrenado; sin embargo, las máquinas tienen aún muchas dificultades para emular capacidades tan humanas como el razonamiento abstracto, la creatividad, el sentido común o la intuición

Para entender mejor el alcance actual de la IA, es necesario diferenciar primero lo que orbita en torno al aprendizaje artificial y lo que no. Y tener claro además que, dentro del ámbito del primero, existen dos grandes familias de técnicas: el aprendizaje supervisado y el no supervi-

sado. El primero es un enfoque que se caracteriza por el uso de conjuntos de datos etiquetados, diseñados para entrenar algoritmos que clasifiquen datos en categorías o predigan resultados con precisión. El segundo analiza y agrupa conjuntos de datos no etiquetados, descubriendo patrones ocultos en los datos sin necesidad de intervención humana. Aunque el aprendizaje no supervisado tiene usos de gran valor (e.g., detección de fraude, aplicaciones genómicas), el momento de esplendor actual de la IA reposa en los resultados obtenidos aplicando los principios de las redes neuronales profundas (llamados así por su gran dimensión con respecto a las existentes desde los años setenta) al problema del aprendizaje supervisado. Sin embargo, el aprendizaje máquina en general, y el supervisado en particular, tiene aún importantes retos por resolver (Hagendorff y Wezel, 2020):

- Muchos algoritmos de aprendizaje son inflexibles: antes de poder utilizar las aplicaciones de aprendizaje máquina (y en particular supervisado), es necesario entrenar los algoritmos correspondientes con una serie de datos relevantes, lo que conduce a una especialización de esos algoritmos. Como consecuencia de lo anterior, el simple hecho de cambiar un color, de modificar una regla en un juego o de omitir letras en los textos, puede dar lugar a un procesamiento automático incorrecto de los objetos o características correspondientes (Rusu *et al.*, 2016).

- Las aplicaciones de IA a menudo carecen de explicación: es bien conocido tanto en la comunidad del aprendizaje máquina como en la industria que aspira

a servirse de estos algoritmos que los algoritmos utilizados se comportan como una caja negra (Pasquale, 2015), a menudo impredecible, que además dista mucho de ser transparente para la trazabilidad de posibles errores y la atribución de las consecuentes responsabilidades. En efecto, las aplicaciones que operan en tiempo real y se reajustan constantemente debido a nuevas entradas pueden producir resultados ininteligibles, y en el caso de operar en entornos de seguridad crítica, producir daños irreparables tanto materiales como humanos.

- Los datos que utilizan los sistemas de IA no siempre representan correctamente la realidad: este desajuste puede aparecer por datos poco fiables, basados en medidas erráticas, falseados o, muy habitualmente, porque, aunque los datos no tengan ninguno de los problemas previos, se almacenen siguiendo pautas con un claro sesgo. Según Northcutt, Athalye y Mueller (2021), los diez conjuntos de datos públicos más utilizados hoy para aprendizaje incluyen un número no despreciable de etiquetas erróneas, lo que está generando un aprendizaje claramente distorsionado. Ese sesgo puede tener diferentes orígenes, que van desde la mala selección de muestras a un volumen inapropiado o descontextualizado, pasado por mecanismos que generan los datos por deducción y no por inducción, dando lugar a información inapropiada.

- El aprendizaje máquina perpetúa el pasado: los algoritmos sobre los que reposan estas técnicas solo pueden operar a partir de los datos existentes, extrapolando patrones apropiados a lo largo de líneas prede-

finidas. Dicho de otro modo, la máquina suele identificar muy bien lo que ya está dado, pero le cuesta reflejar en sus salidas lo que podría o debería ser. Existen contraejemplos aparentes a esta dificultad para adaptarse a nuevos contextos, como la creatividad computacional (Elgammal *et al.*, 2017). Sin embargo, en ese caso, las estrategias adoptadas reposan sobre lo que conocemos como «redes creativas adversas», que se basan en última instancia en patrones, características, estructuras, etc., que ya prevalecen en el mundo. Este fenómeno, además de limitar las capacidades de razonamiento abstracto, es problemático porque en cierto modo perpetúa lo existente (Barocas y Selbst, 2016).

– El pensamiento humano es muy diferente al de una máquina inteligente: las redes neuronales más sofisticadas en la actualidad, con más de mil millones de interconexiones, solo representan un diminuto milímetro cúbico de tejido cerebral. Por otro lado, los métodos estadísticos que sirven de sustrato a las redes neuronales artificiales se alejan en muchos aspectos de los conceptos biológicos de pensamiento, comprensión o aprendizaje. En efecto, cuando se habla de *software* de «reconocimiento» de rostros, sería más preciso decir «conjunto de valores de píxeles que a menudo se correlacionan bien con la presencia de rostros en los datos de entrenamiento que se recogen en el detector» (Elish y Boyd, 2018), porque los ordenadores no «reconocen», «saben» o «aprenden» realmente lo que es un rostro. Esta clara diferencia, que visibiliza las limitaciones de la IA actual, se hace aún más evidente en escenarios específicos, como el de las ilusiones ópticas, absolutamente

indescifrables e irreproducibles para las máquinas. Así pues, tanto la idea de modelar un cerebro a través de *software* como la de replicar estructuras neuronales parecen tener ante sí un largo camino por recorrer.

Todos estos desafíos ponen en evidencia una gran oportunidad de explotar las capacidades actuales, y sobre todo las venideras, en diferentes direcciones. Para entender mejor esas posibilidades, es conveniente explorar el valor del dato en el contexto actual, así como las posibilidades de promover un cambio de paradigma en torno a su uso. Esa posible alternativa buscaría no solo ahuyentar ciertos temores respecto al descontrolado impacto de la IA en nuestro mundo, sino que fomentaría el desarrollo de una nueva economía que, explorando algunos de los retos actuales de la IA, permitiese legitimar su uso desde una perspectiva más sostenible. Las próximas tres secciones aspiran a dibujar el escenario actual desde diferentes ángulos, incidiendo en los aspectos más ligados a la aceptación presente y futura de la tecnología por parte del ciudadano, para retomar después alguna de las áreas de mejora enunciadas en esta sección desde un enfoque propositivo de actuación transversal.

IV. ECONOMÍA DE LO INTANGIBLE

El incesante crecimiento de los datos de los que disponemos –habilitado por el uso masivo de aplicaciones en la nube y del Internet de las Cosas– ha permitido al aprendizaje máquina en general, y al *aprendizaje profundo* en particular, aportar un gran valor tanto a los consumidores, a través

de numerosos servicios, como a las nuevas formas de negocio asociado a ese tráfico de datos. Y lo ha hecho de una forma nunca vista antes, tanto en forma como en intensidad: es, sin duda, un ingrediente fundamental de lo que conocemos como la economía de lo intangible.

Tal como recoge el informe COTEC (2019), los activos intangibles son característicos de la economía del conocimiento e invertir en ellos contribuye a aumentar la productividad y la calidad de vida. Sin embargo, aún no están convenientemente recogidos ni en las estadísticas macroeconómicas oficiales, ni a nivel microeconómico en los libros empresariales. Si se analiza la evolución de la economía mundial en los últimos cuarenta años, parece claro que, aunque la pobreza tiende a reducirse, la desigualdad no para de crecer (Mazzucato, 2021). Las razones de este fenómeno son múltiples y se pueden explicar desde diferentes perspectivas; una de las más utilizadas tiene que ver con la robotización de muchos puestos de trabajo y su impacto (aparentemente) negativo en el empleo. Este aspecto se comentará más en detalle en la sección siguiente, pero sí que es importante destacar aquí una tendencia innegable hacia la especialización, generando brechas importantes entre los empleos que demandan de fuertes competencias en IA y aquellos de bajo valor añadido. Esta no parece que sea la única fuerza que esté generando los desequilibrios mundiales que se observan, pero sí puede ser consecuencia de una dinámica de inversión asimétrica en la I+D, en el *software* o en otros activos inmateriales ligados a la propiedad intelectual (Haskel y Westlake, 2017).

Las diez empresas con mayor facturación del mundo en 1975 fabricaban bienes de servicio –principalmente automóviles– o extraían petróleo. Hoy ninguna de las diez corporaciones más grandes produce bienes tangibles, y la inmensa mayoría de ellas son gigantes tecnológicos que han desarrollado servicios ultrapresentes en nuestras vidas. Estos servicios parecen gratuitos, pero en realidad se alimentan de nuestros datos que, en mayor o menor medida, cedemos como contraprestación a esos servicios. De ahí que muchos califiquen los datos como el petróleo del siglo XXI.

Los bienes materiales, protagonistas absolutos de la economía hasta finales del siglo XX, han dado paso a una economía en la que ya más de la mitad del PIB mundial se genera con activos intangibles. Sin embargo, lo hace de manera muy asimétrica por sectores y áreas geográficas: incluso dentro de la OCDE tenemos países con un 35 por 100 de intangibles, como España, y otros que se acercan al 60 por 100 (países del norte de Europa y EE. UU., COTEC, 2019). Esa brecha, ya de por sí muy importante entre países del primer mundo, es significativamente más grande con respecto a las zonas con menos nivel de desarrollo. La tendencia actual apunta al crecimiento imparable de un grupo de empresas cuyo motor son los datos y la IA, mientras que muchos otros sectores, menos digitalizados, sufren cada día más por mantener su *statu quo*. Como consecuencia de ello: i) existen barreras de entrada cada vez más duras a los sectores tecnológicos basados en IA; y ii) se acentúan, además, las desigualdades de oportunidades de crecimiento entre las ciudades y el medio rural.

El posible reequilibrio de la tendencia en la que estamos inmersos pasa por aproximaciones alternativas, por técnicas de IA aún inexploradas que den respuesta a alguno de los retos planteados en la sección anterior. Sin embargo, para que estas sean competitivas en esta economía de lo intangible se requiere un recurso cada vez más escaso: tiempo. En efecto, es muy difícil invertir cinco o seis años en una tecnología disruptiva sin ayudas o un músculo fuera de lo común (como ha tenido por ejemplo Waymo, la empresa de Google dedicada al desarrollo del vehículo autónomo desde 2012). Solo desde una estrategia pública decidida y a largo plazo, que afronte los riesgos de un cambio de paradigma, se podrá «nivelar» el uso y explotación de la IA. Por desgracia, esa necesidad está a menudo reñida con la visión cada vez más cortoplacista de los que toman decisiones desde lo público, incentivados únicamente por el rédito en intervalos nunca superiores a los ciclos electorales.

Recientemente, Fynn Kidland, Premio Nobel de Economía, decía en relación a las posibles analogías entre los efectos de la COVID en la economía mundial y la que podría generar una guerra: «Hay una gran diferencia: no hay que reconstruir infraestructuras como tras las guerras, pero sí que es más evidente que nunca que hay que proteger el conocimiento humano». La emergencia de un mundo cada vez más digital, con una economía cada vez más orientada a lo intangible y a lo circular, convierte a la IA en un objetivo clave sobre el que se puede y se debe orientar un esfuerzo inversor público que se salga del molde del control presupuestario estándar,

del mismo modo que la pandemia de la COVID-19 ha generado una inyección de liquidez sin precedentes. Una IA alternativa puede dar respuesta al anhelo de Kidland y convertirse en una herramienta esencial para poder dedicar más tiempo a las buenas ideas, a la I+D, para poder aspirar a un paradigma en el que no se deje de lado al ciudadano, sino que se le posicione en el centro. Ese rol protagonista del humano se alcanzará solo con un cambio de estrategia que garantice la transparencia y alimente la confianza, sin penalizar la competitividad. Una estrategia en la que el uso apropiado de los datos será fundamental, y para alcanzarlo la educación será un pilar básico.

V. HACIA UNA GOBERNANZA EQUILIBRADA DE LOS DATOS: EL ROL ESENCIAL DE LA EDUCACIÓN

La recopilación, el uso y el abuso de los datos utilizados para entrenar y alimentar los algoritmos de aprendizaje máquina pueden exponer a las personas a riesgos de los que no son ni conscientes. En efecto, los humanos tenemos grandes dificultades para evaluar correctamente los riesgos de servicios que utilizan IA porque los utilizamos sin pensar todos los días, compartiendo historias, enviando correos electrónicos, buscando información, hablando con máquinas o pidiendo rutas óptimas para el coche.

Para minimizar esos problemas, existen varios frentes de actuación. Parker, Shandro y Cullen (2017) señalan algunas de las buenas prácticas ligadas

a la recogida y al uso de datos: i) la transparencia en cuanto al uso de los datos; ii) la minimización de su uso o lo que se conoce como «privacidad desde el diseño»; y iii) la regulación del almacenamiento y tratamiento de los datos.

Pero un buen uso de los datos desde la perspectiva del usuario no garantiza que las aplicaciones de IA a las que nutren no generen resultados inapropiados. De hecho, existen numerosos ejemplos que ponen en evidencia lo delicado de usar datos con sesgo o incompletos; estas situaciones parecían hasta hace no mucho anomalías estudiadas solo en ambientes académicos; sin embargo, so pretexto de que estas aplicaciones toman mejores decisiones que el humano, su uso cada vez está más extendido en la industria y a nivel gubernamental. Como se explicó en el capítulo anterior, hoy es todavía un reto diseñar mecanismos que estén a la altura de los beneficios prometidos y que, al mismo tiempo, garanticen que sus resultados no tengan sesgo o resulten perjudiciales para la ciudadanía. Y para complicar aún más el desafío, surge la duda de si, una vez conseguido lo anterior, se estará proporcionando una respuesta justa a cada problema, algo con difícil respuesta universal incluso para los juristas. O dicho de otro modo, surge la duda legítima de estar perpetuando o incluso intensificando los males de la sociedad.

Ante semejante perspectiva, la única solución (Borenstein y Howard, 2021) pasa por una reeducación desde las primeras etapas de la interacción con la IA, independientemente de que se sea desarrollador o usuario. La oportunidad de aprender sobre

cómo se utilizan los datos para entrenar la IA, las aplicaciones que la IA puede permitir, etc., debería estar al alcance de cualquier persona que interactúe en cualquier fase con la IA. Tal como sugieren diferentes documentos estratégicos —e.g., Declaración de Montreal (Gibert, Christophe y Guillaume, 2018), informe de IEEE (Chatila y Havens, 2019)— la ética para la IA no debe ser un componente añadido *a posteriori* o una segunda reflexión; debe ser una parte integral en cada etapa del aprendizaje de la IA. En efecto, aunque la mayoría de las personas no diseñarían intencionadamente algoritmos maliciosos, es necesario que se pongan en cuestión sistemáticamente los algoritmos y datos utilizados para identificar los sesgos y los valores que, a menudo inconscientemente, pueden estar incorporando a la tecnología. Sobre este punto se insistirá en la parte final del artículo.

VI. LA AMENAZA DEL EMPLEO Y LA PERSONALIDAD JURÍDICA DE LOS ROBOTS

Tal como se vio en la introducción, es difícil encontrar una definición completa y precisa de la IA. Existe así una nebulosa de capacidades, difícil de acotar, que lleva a menudo a la opinión pública a inferir que la IA es solo aquello que aún no se ha hecho. En efecto, se genera la falsa percepción de que, a medida que las máquinas se hacen más capaces, las tareas que requieren «inteligencia» han de ser más complejas y alejadas de nuestra rutina. Un ejemplo típico de este fenómeno es el reconocimiento óptico de caracteres, que se percibe frecuentemente fuera de

la IA por haberse convertido en una tecnología omnipresente en nuestro día a día.

Con objeto de entender mejor el impacto de la IA en nuestro mundo es más apropiado pensar en una taxonomía que considera dos categorías principales: IA débil e IA fuerte. La primera solo puede tratar con los datos para los que está entrenada y no puede generalizar, acumular experiencia o ser sensible al contexto. La segunda sería aquella que da respuesta a todos los puntos débiles de la primera. Apoyados en los progresos en *aprendizaje profundo* que se describieron anteriormente, ya conseguimos desde hace unos años diseñar algoritmos que pueden detectar ciertas enfermedades mejor que los médicos (Litjens *et al.*, 2015). Y resultados como ese pueden hacer creer en un advenimiento cercano de superinteligencias con motivaciones o prioridades diferentes a las nuestras. Sin embargo, la IA fuerte no es hoy más que una construcción humana, reforzada a través del imaginario colectivo con las películas y novelas de ciencia-ficción. Con los paradigmas en desarrollo, no es algo de lo que tengamos que tener miedo; sin embargo, sí que es cierto que las máquinas pueden hacer tareas, basadas en ciertos patrones aprendidos, cada vez más complejas. Y hacerlo sin cansarse, sin distracciones y sin el factor emocional que tantas veces nos juega malas pasadas al humano. Existe, por tanto, un riesgo serio de que, con el despliegue masivo de la IA, el número de empleos que conocemos hoy se vea significativamente reducido, e incluso de que, en términos netos, tras la incorporación de nuevas ocupaciones hoy en desarrollo, el balance sea negativo.

Sin embargo, esa amenaza no es intrínseca al desarrollo de las tecnologías que embarcan los robots, físicos o virtuales, sino a una lógica empresarial, auspiciada desde los poderes públicos, que se rige por un capitalismo «desestructurado», sin un propósito colectivo. Kate Darling (2021) usa la analogía de nuestra relación con los animales para ilustrar otros paradigmas posibles: «Las diferentes formas en que hemos aprovechado las habilidades de los animales en el pasado demuestran que podríamos optar por diseñar y utilizar estas tecnologías como complemento al trabajo humano, en lugar de intentar simplemente automatizar a las personas». Esta visión entronca con la visión que se tiene en Europa de la IA, tal como se desgranará en la siguiente sección, pero requiere de cambios profundos de la situación actual, tanto estratégicos como regulatorios.

Los sistemas de inteligencia artificial solo se impondrán si se utilizan para fines socialmente aceptados. Los sistemas inteligentes artificiales no tendrán éxito, especialmente en áreas o situaciones en las que las interacciones solo pueden tener lugar de forma adecuada, si no se tienen en cuenta los matices emocionales, la empatía y las sensibilidades subjetivas. En efecto, aunque hay robots que se mueven de forma extremadamente parecida a la humana, hasta el punto de llegar generar el rechazo del famoso «valle misterioso» (Mori, Macdorman y Kageki, 2012), es presumible que poca gente quiera ver un partido de fútbol o una obra de teatro interpretados por robots. Así, al igual que hacemos con los animales, habrá diferentes tipos de relaciones hombre-máquina

inspiradas en el tipo de solución y su contexto de aplicación: para algunas habrá incluso que replantearse seriamente la personalidad jurídica que el Parlamento Europeo comenzó a poner sobre la mesa en 2017 (Bertolini, 2020). Pero para otras, esos derechos y deberes puede que no representen más que una traba en un despliegue que no tiene por que comprometer ningún derecho fundamental de los ciudadanos. Se tratará, pues, de encontrar un equilibrio de geometría variable entre: i) tener un marco de garantías que evite el rechazo social de la tecnología; y ii) no frenar nuestra competitividad, eludiendo a término una pérdida significativa de empleo, que, en todo caso, debería ser cada vez más cualificado.

VII. ¿QUIÉN DIRIGE LA IA Y HACIA DÓNDE?

Dado el vasto alcance que se ha descrito para la IA y sus aplicaciones, ya circulan proyecciones que cifran su impacto potencial en la economía mundial en el orden de los 15.000 trillones de dólares para 2030 (Körner, 2020). Estas previsiones tienen una alta componente especulativa, debido a las fuertes interacciones con otros factores, con variables desconocidas y, sobre todo, a la falta de referencias históricas similares. Pero de cumplirse, y por poner la cifra en perspectiva, estaríamos hablando de un volumen equivalente al de las economías actuales china o de la Unión Europea.

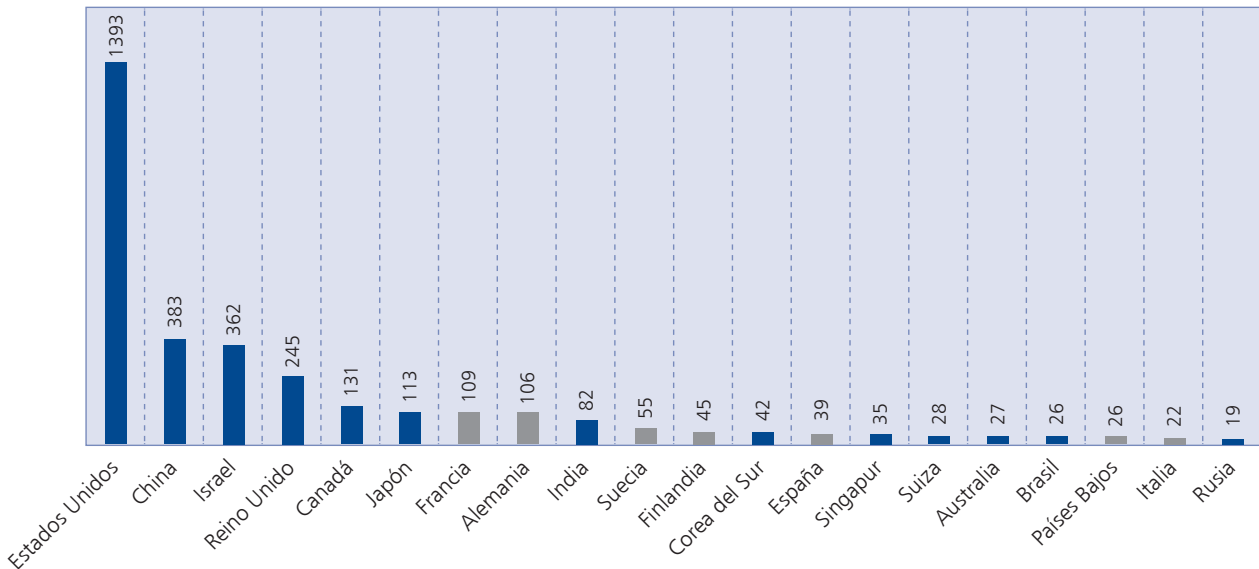
Semejante espiral de negocio ha motivado la aparición en los últimos tiempos de actores inexistentes hace treinta años, que acaparan la práctica totalidad del pastel. Los conocidos Big 9

(Webb, 2019) –Amazon, Google, Facebook, Tencent, Baidu, Alibaba, Microsoft, IBM y Apple) son seis empresas norteamericanas y tres chinas que han conseguido una posición dominante en un mercado que crece exponencialmente y que se apoya de manera decisiva en la inteligencia artificial. Todos ellos se nutren de los datos masivos que les proporcionamos y, como se esbozó en la sección anterior, cavan día a día brechas mayores entre regiones y áreas de negocio, explotando la escalabilidad de lo intangible.

En este contexto, la UE, que sigue siendo la segunda economía del mundo según algunas métricas relacionadas con la IA, se encuentra cada vez más marginada, con un alto riesgo de perder competitividad y riqueza futura. Y es que la posición de esos grandes gigantes no solo afecta a sus cuentas de resultados, sino que estimulan los ecosistemas de innovación y la creación de *start-ups* nativas en IA de sus entornos, favorecidas a menudo por políticas públicas que buscan explotar al máximo el empuje de sus mayores (ver gráfico 1).

Para lograr impulsar la adopción de la IA en los Estados miembros de la UE y seguir siendo competitivos con los gigantes tecnológicos es necesario definir una visión a largo plazo de lo que caracteriza la aproximación europea a la IA. Solo a partir de un planteamiento innovador y de envergadura, se podrán articular políticas ambiciosas que fomenten una cooperación público-privada mucho más arraigada que la actual, que favorezcan la aceptación social para hacer al ciudadano partícipe de este cambio y que establezcan unas reglas del juego claras y consensuadas, pero adaptadas a cada contexto.

GRÁFICO 1
NÚMERO DE *START-UPS* EN TORNO A LA AI, SEGÚN (BERGER, 2018)



VIII. LA VISIÓN DE EUROPA: EL HUMANO EN EL LAZO

A pesar del enorme potencial descrito de la IA en secciones previas, personajes tan influyentes como Barack Obama, Stephen Hawking e incluso Elon Musk han advertido del peligro de una explosión de la IA descontrolada. Siguiendo esa estela de recelos, la ciudadanía comienza a aunar la percepción negativa de una tecnología que les pueda dejar al margen del mercado de trabajo con los miedos a ese descontrol global que evocan los grandes líderes, de las que los Big 9 son los máximos exponentes.

Con el fin de minimizar los efectos adversos y de aumentar su competitividad en declive, Europa tiene aún una oportunidad de usar y mejorar esta poderosísima tecnología desde una perspectiva alternativa a la que vienen marcando las tendencias.

La Unión Europea (UE) cuenta con excelentes centros de investigación, empresas emergentes innovadoras, una posición de liderazgo mundial en robótica y sectores de fabricación y servicios competitivos. Además, ha desarrollado a lo largo de los últimos años una sólida infraestructura informática esencial para el funcionamiento de la IA. Y posee también grandes volúmenes de datos públicos e industriales, cuyo potencial está actualmente infrutilizado (Körner, 2020).

Según propone el *Libro blanco sobre inteligencia artificial* (EC, 2020), Europa debería aprovechar sus puntos fuertes para mejorar su posición en los ecosistemas en los que tiene un rol destacado, creando valor tanto en la fabricación de *hardware* como en el diseño de *software*, pasando por los servicios. A modo de ejemplo, Europa produce más de una cuarta parte de todos los robots industriales y

de servicios profesionales, desempeña un papel importante en el desarrollo y el uso de aplicaciones de *software* para empresas y organizaciones (como la planificación de recursos empresariales, el *software* de diseño e ingeniería). Pero, por encima de eso tiene desde hace no mucho una estrategia (EC, 2020) en la que se busca como en ninguna otra región del mundo el equilibrio adecuado entre regulación, aceptación y la competitividad. Y aspira a hacerlo desde una perspectiva en la que el hombre esté en el centro y en la que la sostenibilidad es un pilar esencial.

El marco reglamentario propuesto debe ser eficaz para alcanzar sus objetivos, sin ser excesivamente prescriptivo, de modo que pueda crear una carga desproporcionada, especialmente dolorosa en las pymes. Al tratar de aplicar esta visión a la gobernanza del dato, la UE debería encontrar formas de superar los

silos existentes y dar a los usuarios plena soberanía para comercializar sus datos de forma segura y sin barreras. Para lograr este equilibrio, la Comisión Europea considera que se debe seguir un enfoque basado en el riesgo de la aplicación de IA considerada, alto riesgo serían por ejemplos sistemas de conducción autónoma y bajo riesgo una aplicación de recomendación de contenidos culturales.

El Reglamento general de protección de datos (RGPD), perfectamente alineado con las buenas prácticas enunciadas en secciones previas, es una herramienta esencial, aunque no suficiente, para dar forma a la aproximación europea de la IA. Es necesario aplicar la regulación que describe a nivel de aplicación/comercialización a todos los proveedores de la UE y de fuera de la UE. Según propone el citado libro blanco, los aspectos del RGPD que podrían afectar al desarrollo de la IA en la UE deben aplicarse estrictamente a los desarrolladores de fuera de la UE que prestan servicios en la UE, tanto para los sistemas de IA de alto riesgo como para los de menor riesgo.

Esta estrategia tiene dificultades de implementación y el tiempo dirá si es viable y competitiva, pero tal como propone un informe del Deutsche Bank (Körner, 2020), la máxima prioridad de la UE debería ser convertirse en una «superestrella» de la IA en consonancia con sus valores, su ordenamiento jurídico y otras prioridades clave, como la política climática y la protección de su orden democrático. Para articular dicha apuesta, la Comisión Europea (EC, 2020) propone caminar hacia una IA de la sostenibilidad invirtiendo en torno a

dos tipos de ecosistemas: i) los de excelencia –respaldando más a las pymes innovadoras y favoreciendo partenariados más potentes entre el sector privado y el sector público–; y ii) los de confianza, que aspiran a dar respaldo a muchas de las preocupaciones y retos esbozados en secciones previas:

– *El hombre en el centro.* Un ejemplo muy ilustrativo de esta idea es la automatización de vehículos. Contrariamente a lo que le llega desde hace años al ciudadano, el vehículo completamente autónomo en cualquier entorno de operación, incluyendo los centros de las ciudades, es una realidad aún lejana. Las razones son múltiples (Villagra et al., 2018; Chai, Nie y Becker, 2021) y no son el foco de este artículo, pero sí es relevante entender la visión que cala cada vez más en el sector: un vehículo con un grado de colaboración humano-máquina variable en función del contexto, pudiendo el conductor descansar en autopista, pero requiriendo su supervisión en entornos más complejos. Este concepto de control compartido, de sinergia entre hombre y máquina, es extensible a multitud de ámbitos, y no solo sería más aceptable por el cerebro humano y su rechazo a la pérdida completa de control, si no que puede resultar incluso más competitivo (se aprovecha lo mejor de cada uno (Acemoglu y Restrepo, 2020)).

– *Seguridad y robustez técnica.* Casi toda la IA que consumimos hoy y la que llegará en los próximos años/meses reposa en servicios que conectan un dispositivo personal con la nube. Esa comunicación bidireccional nube/dispositivo es siempre susceptible de ser atacada y, por tanto, la ciberseguridad se

convierte en un elemento crítico para mejorar la aceptación de estas soluciones. Además, cuando los sistemas en juego son de alto riesgo (e.g., sistemas llamados de seguridad alta o de seguridad crítica como los que embarca un coche o un avión), esa seguridad no solo puede comprometer datos privados sensibles, sino la integridad física de sus usuarios. Hay mucha inversión en juego para mejorar la seguridad actual (*safety* y *security*), pero hay que repensar muchos de esos sistemas para combinar correctamente IA, robustez y predictibilidad.

– *Gobernanza de los datos y privacidad.* El RGPD, en aplicación desde 2018, fue la respuesta de la UE a los temores de vulneración de ciertos derechos de sus ciudadanos y se diseñó para proteger su privacidad y darles más control sobre sus datos personales. Aunque sigue creando controversias sobre la merma de competitividad que puede implicar su aplicación estricta (Pedreño y Moreno, 2020), está perfectamente alineado con el principio de la confianza que necesita el ciudadano, y además ha generado en los usuarios una mayor toma de conciencia de los riesgos en la cesión de los datos.

– *Diversidad, no discriminación y equidad.* La dependencia actual de las técnicas de *aprendizaje profundo* en bases de datos mal etiquetadas, desequilibradas y que incluyeron en su construcción importantes sesgos empieza a generar resultados perniciosos en los algoritmos que alimentan (e.g., herramientas de selección de personal o de identificación de delincuentes). Es necesario, por tanto, no solo ser mucho más exigente con los datos que utilizan los algoritmos que nos

dan servicio, sino también explorar nuevos paradigmas de aprendizaje, en los que los resultados encuentren una traza con el conjunto de datos sobre los que se apoya.

– *Transparencia y accountability (responsabilidad)*. Aunque la corriente de código y datos abiertos cada vez tiene más fuerza entre desarrolladores e incluso empresas, la transparencia necesaria para una IA ética necesita ir mucho más allá. Se ha de documentar completamente el proceso de ingeniería de *software*, los datos, la formación, el rendimiento del sistema, pero además deben surgir nuevas herramientas que ayuden al humano a la inspección sistemática y cuasi-automática de los algoritmos, consiguiendo así procedimientos claros e inequívocos en términos de rendición de cuentas y de responsabilidad.

– *Bienestar social y ambiental*. La tecnología históricamente ha conseguido sus mayores logros bajo el paraguas de grandes retos, a menudo como «subproducto» que no fue planificado inicialmente (e.g., el GPS o la WWW surgieron como necesidad en misiones de mucho mayor calado [Mazzucato, 2021]). Todas las nuevas estrategias y enfoques han de estar alimentados por propósitos concretos, que orienten la actividad de múltiples ecosistemas entre los que, a través de interacciones singulares, surjan aproximaciones disruptivas de la IA. En el caso de la UE, esos propósitos están articulados en torno a los objetivos de desarrollo sostenible, y en cierto modo «aterizados» con el *Green Deal* (Siddi, 2020).

Este singular enfoque puede aplicarse a cualquier ámbito en

el que la IA puede aportar valor, pero existen áreas en las que el *mainstream* en torno al *aprendizaje profundo* todavía no ha llegado con fuerza por su impredecibilidad en contextos de alto riesgo. Europa debe apostar por sectores estratégicos en los que además tiene una gran tradición, como el de la automoción, la salud o la energía, y aportar valor desde una nueva perspectiva en la que primen seguridad, robustez y transparencia. También encajan muy bien con esa visión soluciones de IA centradas en el ser humano, como los sistemas de recomendación híbridos, con responsabilidades compartidas entre el hombre y la máquina. En ambos casos, se debe aprovechar el cambio de paradigma que está en marcha que desplaza la omnipresente computación en la nube por soluciones mixtas en las que la computación *edge* –ocurre en una ubicación física cercana al usuario o la fuente de datos– o incluso *fog* –nodos interconectados de procesamiento previos a la nube– cada vez tienen más peso. Y en paralelo, apoyar sin titubeos el desarrollo de dos nuevas tecnologías en ciernes, como la computación neuromórfica y la computación cuántica, que pueden contribuir decisivamente a encontrar una ventaja competitiva en este campo.

IX. EVOLUCIÓN O REVOLUCIÓN

La inacción o el diseño de políticas tímidas no es una opción. El dinamismo actual no permite ser conservador. Se ha demostrado que el cambio de tendencia es aún posible: queda mucho por hacer y, sobre todo, por hacer de otra manera. Y, además, el mundo se enfrenta a

grandes retos asociados a la sostenibilidad. Pero para articular un cambio de tendencia hay que actuar con vigor en diferentes frentes. La estrategia planteada por Europa en la sección anterior implica, sin duda, un cambio importante con respecto a las que siguen en EE. UU. y China, así que es inevitable tener dudas sobre su idoneidad. Esas dudas serán tanto más grandes cuando menos clara esté la estrategia para la ciudadanía y menos ambiciosas y efectivas sean las políticas puestas en marcha. Y para articular esas políticas, será clave tener en cuenta tanto el cómo como el para qué. Siguiendo los consejos de Mazzucato (2021), los poderes públicos deben cambiar su aproximación. No es viable seguir siendo observadores que entran en escena solo para corregir desequilibrios. La UE, como paraguas político, y los Gobiernos de los Estados miembros como conocedores de las especificidades de cada territorio, tienen que emprender una transformación, orientando sus pasos hacia organizaciones innovadoras, con capacidad de arriesgar para catalizar la actividad hacia sectores y proyectos cuyo propósito sea una IA más sostenible, segura y centrada en la persona, capaz de resolver los problemas que acucian a nuestra sociedad. La revolución necesaria ha de afrontarse desde diferentes puntos de vista, pero se plantean a continuación tres ejes ineludibles:

– Es capital hacer sentir a los ciudadanos partícipes del nuevo mundo que viene, en el que la IA aparece en todos los niveles, e implicarles en el cambio de enfoque respecto a la tendencia actual. Cómo educar a los niños y los ciudadanos en la IA es clave: i) para la aceptación de la tecno-

logía; y ii) para hacerles protagonistas del cambio que traen las tecnologías inteligentes desde un punto de vista a la vez crítico y constructivo. En este sentido, es esencial profundizar en la Declaración de Pekín de mayo de 2019 sobre la IA y Educación, firmada por más de cien Estados, que permitió llegar a un consenso sobre la urgencia de que los Gobiernos capaciten a sus ciudadanos para poder adaptarse a los cambios requeridos por una implantación generalizada de sistemas de IA. La declaración también desea promover el uso equitativo e inclusivo de la IA, independientemente de potenciales discapacidades, del estatus social o económico, de antecedentes étnicos o culturales, o de ubicaciones geográficas. Enfatiza además la igualdad de género, y tiene como objetivo garantizar el uso ético, transparente y auditable de los datos de o para la educación, entre otras recomendaciones. Esta nueva mentalidad debería llegar a todas las organizaciones que desarrollen o usen IA para consolidar en el imaginario colectivo que cualquier toma de decisiones, incluidas las técnicas en el diseño de IA, está entrelazada con consideraciones éticas.

– Deben asumirse riesgos desde los poderes públicos, que han de recuperar el liderazgo, conformando y no solo corrigiendo, desarrollando capacidades y competencias para una IA orientada por misiones. Los riesgos y recompensas han de compartirse de manera más equitativa (sector público/privado), siempre en torno a retos que, además de ambiciosos, han de ser inspiradores para todos. Así, no solo el investigador o desarrollador ha de sentirse motivado por el desarrollo de un nuevo algoritmo

de IA, sino que el conjunto de la ciudadanía ha de compartir el propósito de lo que se persigue, de manera similar a la comunidad surgida con la comunidad científica en el último año para el desarrollo de las vacunas contra el virus SARS-CoV-2. En esta línea, es también esencial que las administraciones públicas sean partícipes de ese nuevo enfoque y comiencen rápidamente a utilizar y desplegar productos y servicios que se basen en la IA para sus actividades.

– Fomentar con políticas de incentivos –sistemas de precios o ayudas públicas– el pensamiento creativo, la interdisciplinariedad y la asociación público-privada de amplio espectro. Para ello, hay que incidir en las iniciativas actuales, que usan como instrumentos fundamentales los *Digital Innovation Hubs*, plataformas de IA a la carta (EC, 2020) o redes paneuropeas de investigación (e.g., Ellis, 2021). Además, se debe fortalecer el rol de las pymes en este proceso innovador. En una visión inspirada por grandes misiones, todas las tecnologías que aparezcan por el camino podrán ser relevantes. Este efecto derrame, del que hay numerosos ejemplos en grandes proyectos del pasado, tiene en la economía de lo intangible un efecto si cabe más acusado, y las pequeñas estructuras innovadoras son actores fundamentales para capitalizar esas oportunidades, a menudo imprevistas. Hay que apoyar mucho más a los *start-ups*, pero no solo por sus capacidades tecnológicas en torno a la IA o la robótica, si no, sobre todo, por su capacidad para aportar valor a una gran misión, interactuando de la mayor manera posible con ecosistemas sinérgicos y con el ciudadano.

Y en esta foto general de cómo debe afrontar Europa el inmenso reto de su competitividad en torno a la IA, surge la pregunta de cómo debe prepararse España para los tiempos venideros. Como Estado miembro influyente, debe, por supuesto, seguir la estela de esta estrategia ambiciosa, pero ha de buscar su singularidad y valor diferencial intentando contribuir a resolver algunos de los problemas globales que más le acucian, como, por ejemplo: i) la lucha contra el cambio climático con la sustitución de combustibles fósiles por renovables; ii) el diseño e implementación de ciudades inteligentes y climáticamente neutras con vistas, entre otros, a un turismo sostenible; y iii) la mejora y conservación de la salud del suelo y de los alimentos.

Esta visión y sus políticas asociadas –esbozadas parcialmente en la Estrategia Nacional de Inteligencia artificial (ENIA, 2020)– tienen como objetivo conseguir un cambio de tendencia que en España es, si cabe, mas necesario. En efecto, es necesario incrementar el peso específico de la IA en el mundo empresarial; es aún raro ver organizaciones con una visión unificada en torno a la IA, hay aún mucho desconocimiento para evaluar proveedores correctamente o para integrar cambios de calado en las estructuras y procedimientos actuales y, además, el entramado legal dificulta en ocasiones el despliegue masivo de soluciones. En el panorama actual se observa aún una importante desconexión entre lo que se puede hacer y lo que realmente se está haciendo, en particular en pymes no nativas digitales. Esta falta de preparación es especialmente crítica en los entornos productivos en los que es cada vez más necesario

un enfoque orientado a Industria 4.0 (dominada por una creciente digitalización y los sistemas ciberfísicos) e incluso una transformación al paradigma Sociedad 5.0 (por el que se aspira a una producción ultrapersonalizada, aprovechando que la generación de conocimiento e inteligencia la hacen las propias máquinas al servicio del usuario/consumidor). Por todo ello, la difusión, la formación y la educación son una piedra angular para la integración satisfactoria de la IA en nuestras vidas y en consecuencia para una mayor competitividad de la economía española. Y esta línea de actuación ha de afectar a las políticas públicas, al desarrollo reglamentario, a las hojas de ruta de financiación de la I+D y a las propias estrategias de las empresas. Ya no es una opción seguir la estela de otros, es necesario recorrer cuanto antes ese camino diferencial que nos permita ser competitivos en Europa y en el mundo.

La ENIA está bien orientada para seguir ese camino. Propone seis ejes de actuación que tiene conexiones importantes con los planteamientos anteriormente descritos: i) impulsar la I+D+i en IA; ii) potenciar y atraer talento; iii) desarrollar infraestructuras relevantes que den soporte a la IA; iv) integrar la IA en las cadenas de valor actuales; v) potenciar el uso de la IA en la Administración Pública y en misiones estratégicas nacionales; y vi) establecer un marco ético y normativo alineado con la visión europea. Se propone aplicar esas acciones para dar respuesta a desafíos sociales como la reducción de la brecha de género y digital o el favorecimiento de la transición ecológica y la vertebración territorial, elementos de indudable interés y cuyo tratamiento es

necesario. Sin embargo, se echa en falta una mejor definición de esas misiones estratégicas específicas para la economía española en la que la IA puede ser: i) una palanca decisiva para la competitividad; y ii) una fuente de inspiración para otros sectores productivos. La ENIA es un buen punto de partida, pero ha de ser mas ambiciosa y audaz. Una cierta evolución del *statu quo* no parece suficiente, es necesaria una pequeña revolución hacia una robótica e inteligencia artificial más competitivas, humanas y sostenibles.

BIBLIOGRAFÍA

- ACEMOGLU, D. y RESTREPO, P. (2020). The wrong kind of AI? Artificial intelligence and the future of labour demand. *Cambridge Journal of Regions, Economy and Society*, 13(1), p. 25-35.
- BAROCAS, S. y SELBST, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), pp. 671-732.
- BERGER, R. (2018). *Artificial Intelligence-A Strategy for European Startups. Recommendations for policy makers*, Asgard-Human Venture Capital. Berlin.
- BERTOLLINI, A. (2020). *Artificial Intelligence and Civil Liability*. European Parliament.
- BORENSTEIN, J. y HOWARD, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, 1(1), pp. 61-65.
- CHAI, Z., NIE, T. y BECKER, J. (2021). Top Ten Challenges Facing Autonomous Driving. En *Autonomous Driving Changes the Future*, pp. 137-178. Springer.
- CHATILA, R. y HAVENS, J. C. (2019). The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. En ALDINHAS FERREIRA, M., SILVA SEQUEIRA, J., SINGH VIRK, G., TOKHI, M., E. KADAR E. (eds.), *Robotics and Well-Being. Intelligent Systems, Control and Automation: Science and Engineering*, vol 95. Springer,

Cham. https://doi.org/10.1007/978-3-030-12524-0_2

- COTEC (2019). *Innovación en España: Informe COTEC 2019*. Fundación COTEC para la Innovación.
- DARLING, K. (2021). *The New Breed: What Our History with Animals Reveals, about Our Future with Machines*. Henry Holt.
- EC (2020). *White Paper on Artificial Intelligence – A European approach to excellence and trust, COM (2020)*. European Commission.
- AHMED ELGAMMAL, A., LIU, B., ELHOSEINY, M. y MAZZONE, M. (2017). CAN: Creative adversarial networks, generating art by learning about styles and deviating from style norms., *arXiv preprint arXiv:1706.07068*.
- ELISH, M. C. y BOYD, D. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs*, 85(1), pp. 57-80.
- ELLIS (2021). <https://ellis.eu>
- ENIA (2020). <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/021220-ENIA.pdf>
- GIBERT, M., CHRISTOPHE, M. y GUILLAUME, C. (2018). *Montréal declaration of responsible AI: overview of international recommendations for AI Ethics*. University of Montréal.
- HAGENDORFF, T. y WEZEL, K. (2020). 15 challenges for AI: or what AI (currently) can't do. *AI and Society*, 35(2), pp. 355-365.
- HASKEL, J. y WESTLAKE, S. (2017). *Capitalism without capital*. Princeton University Press.
- KÖRNER, K. (2020). How will the EU become an AI superstar?, Deutsche Bank Research.
- LECUN, Y., BENGIO, Y. y HINTON, G. (2015). Deep learning. *Nature*, 521(7553), pp. 436-444.
- LITJENS GEERT, J. S., BARENTSZ, J. O., KARSSEMEIJER, N. y HUISMAN, H. J. (2015). Clinical evaluation of a computer-aided diagnosis system for determining cancer aggressiveness in prostate MRI. *European Radiology*, 25(11), pp. 3187-3199.

<p>MAZZUCATO, M. (2021). <i>Mission economy: A moonshot guide to changing capitalism</i>. Penguin UK.</p> <p>MCCARTHY, J. (1983). The little thoughts of thinking machines. <i>Psychology Today</i>, 17(12), pp. 46-49.</p> <p>MCCARTHY, J. (1998). <i>What is artificial intelligence?</i> Disponible en: http://www-formal.stanford.edu/jmc/</p> <p>MORI, M., MACDORMAN, K. F. y KAGEKI, N. (2012). The uncanny valley (from the field). <i>IEEE Robotics & Automation Magazine</i>, 19(2), pp. 98-100.</p> <p>NORTHCUTT, C. G., ATHALYE, A. y MUELLER, J. (2021). <i>Pervasive label errors in test sets destabilize machine learning benchmarks</i>. arXiv preprint arXiv:2103.14749.</p> <p>PARKER, N., SHANDRO, A. y CULLEN, E. (2017). <i>Autonomous and connected</i></p>	<p><i>vehicles: navigating the legal issues</i>. London: Allen & Overy LLP.</p> <p>PASQUALE, F. (2015). <i>The black box society</i>. Harvard University Press. FIIA Working paper, Finnish Institute of International Affairs.</p> <p>PEDREÑO, A. y MORENO, L. (2020). <i>Prevenir el declive en la era de la inteligencia artificial</i>, Kindle Direct Publishing.</p> <p>PERRAULT, R., et al. (2019). <i>Artificial Intelligence Index Report 2019</i>. Technical report. Stanford Institute for Human-Centered Artificial Intelligence.</p> <p>RAJPURKAR, P., JIA, R. y LIANG, P. (2018). Know what you don't know: Unanswerable questions for SQuAD. <i>Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics</i>, pp. 784-789. Melbourne, Australia, 2018.</p>	<p>RUSU, A. A., et al. (2016). <i>Progressive neural networks</i>. arXiv preprint arXiv:1606.04671.</p> <p>SHANAHAN, M. (2015). <i>The technological singularity</i>. MIT Press.</p> <p>SIDDI, M. (2020). The European Green Deal: Assessing its current state and future implementation. <i>FIIA Working Paper</i>, May 2020.</p> <p>VILLAGRA, J., ACOSTA, L., ARTUÑEDO, A., BLANCO, R., CLAVIJO, M., FERNÁNDEZ, C., GODOY, J., HABER, R., JIMÉNEZ, F., MARTÍNEZ, C., NARANJO, J. E., NAVARRO, P. J., PAÚL, A. y SÁNCHEZ, F. (2018). Automated driving. <i>Intelligent Vehicles: Enabling Technologies and Future Developments 8</i>, pp. 275-342.</p> <p>WEBB, A. (2019). <i>The big nine: How the tech titans and their thinking machines could warp humanity</i>. UK: Hachette.</p>
---	---	---