

**EL SABER EN LA ÉPOCA DE SU REPRODUCTIBILIDAD TÉCNICA.
SOBRE LA RELACIÓN ENTRE LA INTELIGENCIA ARTIFICIAL Y LA
HUMANA**

***THE KNOWLEDGE IN THE AGE OF MECHANICAL REPRODUCTION.
ON THE RELATIONSHIP BETWEEN ARTIFICIAL AND HUMAN
INTELLIGENCE***

JOSÉ LUIS DELGADO ROJO

Istituto Italiano di Scienze Umane - Scuola Normale Superiore,
Italia
jluisdelgado@hotmail.com

RECIBIDO: 20/08/2023

ACEPTADO: 24/11/2023

Resumen: El éxito reciente de la inteligencia artificial supone uno de los mayores desafíos a los que tiene que hacer frente el ser humano en la actualidad, en especial debido a la irrupción de nuevas formas de crear y adquirir conocimiento. El presente trabajo trata de describir estos cambios así como los nuevos tipos de relación que la inteligencia humana puede establecer con la artificial, apostando por la complementariedad y dejando al margen la tesis del control. Finalmente se explora cómo pueden traducirse estas relaciones en el ámbito de la educación, donde la cuestión del conocimiento es central.

Palabras clave: inteligencia artificial; antropología filosófica; educación; conocimiento; control; coevolución

Abstract: The recent success of artificial intelligence represents one of the greatest challenges that humanity has to face today, especially due to the emergence of new ways of creating and acquiring knowledge. This paper aims to describe these changes as well as the new types of relationships that human intelligence can establish with artificial intelligence, focusing on complementarity and leaving aside the thesis of control. Finally, it explores how these relationships can be translated into the field of education, where the question of knowledge is central.

Keywords: artificial intelligence; philosophical anthropology; education; knowledge; control; coevolution

Introducción

El desarrollo de la tecnología más reciente, la inteligencia artificial (IA), obliga a volver sobre algunas de las preguntas más antiguas de la filosofía. La aparición de un alter ego digital, capaz de aparecer como competencia en tareas, como las cognitivas, en las que hasta ahora el ser humano tenía la exclusiva, nos lleva a desempolvar de nuevo algunas de las viejas cuestiones. ¿Qué es el ser humano? ¿En qué consiste su singularidad? ¿Puede hablarse todavía de una diferencia humana? La interrogación sobre nosotros mismos parece haber estado siempre lastrada por un obstáculo insuperable: la pregunta no fructifica en el vacío de la propia interioridad y necesita de un término de contraste externo respecto al cual afirmar o negar nuestra diferencia. Primero fueron los dioses, en los que los griegos, por ejemplo, vieron los términos con los que interpretar algo que ya existía pero no tenía nombre todavía y que se abrió paso lentamente a la observación racional dando lugar a la invención o descubrimiento de la esfera interior del espíritu (Snell, 2007), hoy diríamos, subjetividad. Después, la vida animal ofreció el modelo sobre el que la antropología filosófica ha vuelto neuróticamente una y otra vez como apoyo con el que definir, *a contrario*, la esencia humana. La irrupción de la IA supone una nueva vuelta de tuerca sobre la misma cuestión y pasa a funcionar hoy como “el espejo con cuya ayuda [el ser humano] se ve a sí mismo” (Snell, 2007, p. 340). El ser humano solo se entiende a partir de sus exteriorizaciones. ¿Qué tipo de imagen le devuelve este espejo? ¿Ofrece la IA alguna forma de autoconocimiento o más bien representa la ocasión para nuevas formas de automistificación?

Kant señaló de forma célebre que la pregunta “¿qué es el hombre?” era demasiado difícil de abordar directamente pero equivalía a otras tres preguntas, que una vez contestadas ofrecerían la respuesta a la anterior: ¿qué puedo saber?, ¿qué debo hacer? y

¿qué debo esperar? Vamos a centrarnos en lo que sigue en una estrategia doblemente indirecta para atacar la pregunta temible de la antropología filosófica, planteando la primera pregunta de la serie - “¿qué puedo saber?”- en el actual contexto de disrupción tecnológica e intentando obtener alguna respuesta a partir de lo que da a pensar la transformación en el campo del conocimiento que ha producido la aparición de un competidor cognitivo al ser humano como la IA.

Exteriorización

Vivimos en la época de la inteligencia artificial. Podría parecer entonces que lo que está pasando ahora es un acontecimiento radicalmente nuevo y sin precedentes. ¿Es así realmente? Muchas de las cosas que parecen “realmente nuevas” en realidad forman parte de una tendencia que viene de lejos y que puede ayudar a encontrar puntos de referencia para juzgarlas, disolviendo así el aura de misterio con las que las invisten (vamos a suponer que inconscientemente) tanto sus partidarios como sus detractores. No se trata de posicionarse a favor o en contra sino de algo previo: de orientarse para intentar entender.

La IA es el último avatar de una larga historia de “exteriorización” de capacidades humanas. “Exteriorizar” es el proceso de materializar u objetivar en un soporte externo una capacidad o habilidad que hasta ahora había tenido como soporte exclusivo al ser humano. El origen de esta larga historia se encuentra en la invención de la escritura, dado que escribir supone fijar en una memoria externa una capacidad que hasta el momento estaba unida de forma rígida a estructuras cognitivas humanas (Stiegler, 2002, pp. 192 y ss). Es posible que la escritura fuera precedida por el ritual, como protolenguaje que fija un significado simbólico en significantes externos - la coreografía corporal del ceremonial ritual

(Rappaport, 2016). En todo caso, no importa si es a través del papel, de tablas de cera o con cualquier otro medio más sofisticado, ya que la novedad consiste en traspasar información a un soporte físico que mantiene su integridad a través del paso del tiempo sin nuestra intervención activa.

Esta exteriorización posteriormente alcanza al campo de la percepción, con la invención en el siglo XIX de la cámara fotográfica o el telégrafo, que de nuevo permiten fijar en un soporte externo imágenes o sonidos. Se exterioriza así en un medio técnico la percepción visual o auditiva humana, es decir, capacidades o *competencias humanas*. Otros sentidos, como el tacto o el gusto, parece que se resisten por ahora a la exteriorización, aunque probablemente es solo cuestión de tiempo ya que se está trabajando intensamente desde hace algunas décadas con este objetivo.

La IA se puede describir entonces como la culminación, hasta el momento, de esta tendencia a la exteriorización que acompaña desde el principio a la historia del desarrollo tecnológico. Ahora le ha llegado el turno a la capacidad que parecía una prerrogativa exclusivamente humana: la capacidad cognitiva (el conocimiento, pensamiento o inteligencia), que vemos ahora realizada por un medio o dispositivo externo, el cableado *software* de las redes neuronales. La aparición de la ingeniería cognitiva implica que somos capaces de reproducir técnicamente un proceso o capacidad que hasta ahora se daba únicamente de forma natural (en un soporte biológico). Podemos extraer algunas conclusiones preliminares de este breve repaso histórico.

¿Extensión o exaptación?

Una interpretación no muy sofisticada consistiría en ver la exteriorización como una continua “extensión” de nuestras

capacidades (como decía McLuhan por ejemplo), es decir, una tendencia hacia la progresiva optimización con medios técnicos de capacidades humanas. Esta interpretación de la técnica como prótesis que sustituye a la vez que potencia es una versión apenas encubierta del relato moderno del progreso, donde cada etapa supera las anteriores (en el sentido hegeliano de eliminar lo superfluo y a la vez conservar lo válido en una versión mejorada), siguiendo un movimiento de mejora continuo. Este sigue siendo el relato que hoy en día sostienen los partidarios del transhumanismo. Curiosamente es también la base del relato apocalíptico de sus detractores, para los cuales la tecnología, en ese proceso imparable de mejora de capacidades humanas, hará del ser humano algo irrelevante y finalmente acabará por sustituirlo.

¿Qué tiene de miope esta interpretación? Precisamente que supone que lo que es mejorado no cambia cuando aparece la mejora. La historia de una mejora lineal y continua presupone que el ser humano es una especie de sustrato permanente al margen de los cambios que lo mejoran en otro soporte. Pero ¿qué pasaría si cada mejora lograda en el proceso de exteriorización en realidad no dejara intacto al ser humano sino que de alguna manera retroactuara sobre él y le forzara a modificar sus capacidades para adaptarse al nuevo entorno transformado debido a su intervención? Entonces la “extensión” del ser humano a la vez implicaría *autotransformación* y el relato lineal se curvaría hasta adoptar la forma de un bucle: es cierto que el hombre se exterioriza en sus dispositivos pero a la vez tiene que modificar sus capacidades para adaptarlas a la presión selectiva que le imponen esos mismos dispositivos. A diferencia del relato lineal, en un bucle el siguiente paso es siempre imprevisible: la historia permanece abierta y solo depende del tipo de interacción que puedan establecer los dos polos, el ser humano y sus exteriorizaciones.

No tendríamos que hablar entonces de “extensión” -y por tanto

de progreso- sino de “exaptación” -y por tanto de coevolución hombre-máquina. La “extensión” es la mitad del proceso, que necesita de la otra mitad (la “exaptación”) para cerrar el bucle que permite entender las implicaciones humanas de la aparición de una nueva tecnología.

En biología evolutiva la “exaptación” es el nombre de un proceso que hace referencia al cambio de función de un órgano debido a cambios en el entorno. A diferencia de la “adaptación”, en la que la aparición o permanencia de un órgano se explican por la ventaja que su función ofrece al organismo en su relación con el entorno, la “exaptación” implica el cambio de función de un órgano que ya existía, aunque tuviera otra y, por tanto, la reconfiguración interna que sufre el organismo bajo la presión del medio. Los órganos no son meramente sustituidos por otros sino que se mantienen con funciones diferentes. La adaptación supone la sustitución o superación de lo viejo por lo nuevo, la exaptación supone en cambio un nuevo uso de lo viejo. No hay optimización sino diferencia. La adaptación produce lo mismo pero mejorado, la exaptación produce algo diferente.

Un ejemplo clásico es la descripción que hace Leroi-Gourhan de la bipedestación: cuando el ser humano logró la posición erecta las manos perdieron la función locomotora, que pasó a las piernas, pero a la vez adquirieron una nueva función, la de exploración táctil y prensil, que a su vez liberó al rostro de estas funciones (la prensión con los dientes, por ejemplo) y permitió que ganara otras nuevas: el rostro para expresar emociones, la boca para hablar y los ojos para tener un horizonte visual más amplio, lo que dio lugar a la visión a distancia.

La refuncionalización es un proceso habitual en la historia de la recepción de la técnica. Así, por ejemplo, se ha señalado frecuentemente que la aparición de la escritura (la “técnica de la palabra”) comportó una reconfiguración de las funciones cognitivas

humanas, que una vez liberadas de la necesidad de la memorización literal, ahora delegada en el registro escrito, pudieron adquirir otras nuevas: la exploración imaginativa, la innovación y el pensamiento abstracto (Ong, 2016; véase también Serres, 2014, p. 280). También por lo que respecta a la percepción se puede indicar un ejemplo similar. Cuando aparece una nueva tecnología a menudo la primera reacción consiste en un proceso de imitación, que hace con lo nuevo lo mismo que hasta ahora se hacía con la tecnología anterior. También esto se aplica al caso de la escritura, que en sus inicios “imita por escrito la producción oral”, antes de conseguir independizarse y afirmar las propias potencialidades del nuevo medio (Ong, 2016, p. 67). La aparición de nuevos medios como la fotografía o el cine, así como la aparición de nuevos materiales para la arquitectura, desencadenó al inicio una respuesta mimética: la fotografía, por ejemplo, se utilizaba para hacer retratos, a imitación de la pintura, y el acero se continuó utilizando para los ornamentos florales que antes se hacían con madera. Es solo en un segundo momento cuando se aprecia la potencialidad del nuevo medio artístico y esto produce un cambio en el sentido de la propia actividad artística, que ya no se realizará en los mismos términos que hasta ahora. Walter Benjamin ha aportado multitud de ejemplos en las secciones F y G del Libro de los Pasajes sobre la oposición entre el momento “histórico” y el “constructivo” de la recepción de la técnica en el siglo XIX. El uso *kitsch* de la tecnología del siglo XIX se puede detectar también en las primeras producciones en serie, que se dedican en un primer momento a fabricar de forma industrial copias de productos hechos a mano, desde muebles a los primeros coches, que al principio son producidos unidad a unidad, mientras que solo en un segundo momento aparece una forma de creatividad propiamente industrial (Lipovetsky y Serroy, 2015, pp. 130-133). Por tanto, ser creativo no quiere decir lo mismo antes que después de la aparición de los nuevos medios.

También en el caso de la IA nos podemos hacer la misma pregunta: ¿qué nuevo sentido de nuestras capacidades pueden liberar o hacer surgir en nosotros las nuevas exteriorizaciones técnicas? ¿O tenemos que resignarnos a un uso *kitsch* de la técnica, como el que supone la “extensión” transhumanista de capacidades, que pretende mejorarlas sin cambiarlas, optimizar el órgano sin cambiar su función, como si nuestras capacidades fueran un paquete de rasgos fijos y no pudieran cambiar, por ejemplo, a través de la interacción con las máquinas? Podemos aprender del pasado e intentar evitar algunos de sus errores explorando alternativas a la recepción *kitsch* de la IA, ejemplificada actualmente en la tesis del control (dominante hoy en día en el tratamiento de la relación humano-máquina). Según reza esta tesis, para evitar ser dominados por la IA la tendremos que dominarla nosotros, utilizándola como medio subordinado a nuestros objetivos (un ejemplo representativo se puede encontrar en Russell, 2019 y Tegmark, 2018). Uno de los problemas que plantea la cuestión del control es que se define a partir del sujeto que lo ejerce (control es lo que hacemos para protegernos de cambios súbitos e imprevisibles) y, por tanto, presupone o da por bueno un sujeto con una dotación de capacidades relativamente inalterable y al margen de los cambios. Una forma diferente de relacionarse con la técnica pasará por relativizar la tentación del control y aceptar una cierta apertura e indeterminación en la relación mutua entre el ser humano y sus aparatos.

Inteligencias múltiples

Hemos visto que el sujeto no queda intacto después de la exteriorización. Veremos a continuación que algo parecido le pasa a lo que es exteriorizado. Una segunda conclusión que se puede extraer de la concepción de la técnica como exteriorización es que la

capacidad exteriorizada no supe exactamente a aquello que exterioriza. Por eso la metáfora de la técnica como prótesis (órgano supletorio) sólo capta una parte de la verdad. En realidad la capacidad humana (memoria, percepción o cognición) cambia de naturaleza cuando cambia de soporte, del biológico al inorgánico. Se convierte en otra cosa.

Aceptar la diferencia cualitativa entre la capacidad natural y la artificial permite cuestionar de nuevo tanto la historia de progreso como la de catástrofe. En efecto, por lo que respecta a la primera, la posthumanidad (el hombre potenciado gracias a la tecnología) presupone que la técnica amplifica y por tanto aumenta cuantitativamente una capacidad que mantiene la homogeneidad cualitativa con aquello que es aumentado (ya que sino no habría progreso). Por lo que respecta a la segunda, el miedo a nuestra sustitución por la IA o la presión por competir con ella presupone que las máquinas saben hacer mejor lo que nosotros no hacemos tan bien, que pueden triunfar donde nosotros hemos fracasado. El miedo o la competición son formas de reacción que presuponen, por tanto, que las capacidades técnicas son del mismo tipo que las humanas, con solo una diferencia en el más o el menos (para una crítica de este presupuesto véase Coeckelbergh, 2022, pp. 32-33).

Este supuesto resulta ser muy discutible. Por lo que respecta a la memoria, la humana y la técnica funcionan de forma muy diferente. La primera triunfa en el orden y en ámbitos dotados de estructura pero en cambio es poco precisa con el detalle literal. La segunda es una memoria literal y puede guardar con exactitud un elevado volumen de datos pero en cambio es ciega al orden o desorden. La oposición corresponde a la distinción que Simondon hace entre “plasticidad del soporte” y “plasticidad del contenido” (Simondon, 2008, pp. 138-140). Dado que cada una tiene un alto rendimiento donde el de la otra es bajo, esto sugiere la posibilidad de una relación de “complementariedad”, donde la técnica puede suplir las

insuficiencias humanas pero sin el riesgo de sustituir sus capacidades.

También en el caso de la percepción se ha señalado que el ojo técnico no funciona como el ojo humano: el primero realiza un gesto instantáneo y estático mientras que el segundo se compone de una multiplicidad de microgestos que captan de forma exploratoria la superficie que miran y luego sintetizan todos estos aspectos en un objeto virtual que no tiene correlato físico directo (Virilio, 1998, pp. 9-26).

¿Y en el caso del conocimiento? Si partimos de una definición en principio neutra del conocimiento como “procesamiento de información” parece que sí sería posible equiparar las prestaciones humanas con las técnicas. El reconocimiento de imágenes o la generación de lenguaje natural, tareas en las que la IA está obteniendo un elevado rendimiento, se basan en procesos algorítmicos que lo que hacen es, efectivamente, procesar información. Tal vez el problema esté en que la definición inicial resulta no ser tan neutra como parecía ya que, en virtud de su misma vaguedad, predispone a aceptar una cierta equiparación entre la inteligencia natural y la artificial, cuando en realidad, vistas más de cerca, sus diferencias pasan a un primer plano. El ser humano y la IA no ocupan el mismo “nicho cognitivo” por lo que pueden cohabitar en el mismo ecosistema sin competición e incluso explorar formas de cooperación mutua (Lee, 2020, p. 160). Este enfoque de una diferencia cualitativa entre distintos tipos de inteligencia ayuda a plantear la pregunta correcta sobre la IA, que no es tanto “qué pasará” (cuestión de especulación y futurología) como “qué debemos hacer” ahora en el presente. Es decir, ¿cómo hemos de utilizarla, en función de lo que sabemos sobre su funcionamiento actual y de acuerdo con el estado actual de nuestras capacidades?

Por tanto, si queremos describir el nuevo papel que le corresponde al ser humano como cooperador de las máquinas es

necesario saber qué diferencia el conocimiento humano del artificial. No hay una única forma de inteligencia sino que cada una tiene un rendimiento diferente en diferentes condiciones. Sabiendo los puntos fuertes y débiles de cada una podremos entonces identificar en qué capacidades ha de especializarse el ser humano y en cuáles puede delegar en la máquina para reforzar la relación de complementariedad entre ambos.

En la época de la reproductibilidad técnica de la capacidad cognitiva, el ser humano tiene que identificar la especialidad humana si quiere preservarla y desarrollar su espacio propio. Su tarea será la de especializarse en la propia humanidad. Si se dedica en cambio a copiar e imitar a sus productos entonces la tesis de la sustitución se volverá verdadera de manera performativa: si creemos que hemos de competir en el mismo nicho que las máquinas entonces nos comportaremos de forma que seremos finalmente sustituidos por ellas. Si ponemos en cambio el acento en las diferencias no es para satisfacer imaginariamente una falsa sensación de dominio. Esta idea del ser humano ya no puede sostenerse. No se puede mirar a otro lado para no ver la herida narcisista que supone la aparición de otra forma de inteligencia donde hasta ahora reinaba la nuestra en solitario. Al contrario, identificar la diferencia con las máquinas nos da la clave para saber cómo establecer un nuevo vínculo y acercarnos productivamente a ellas.

¿Cuáles son entonces las diferencias entre los dos aparatos cognitivos? ¿Qué quiere decir conocer, en nuestro caso y en el suyo? ¿Pueden las dos formas coexistir y, es más, cooperar?

El fin de las competencias

El antropólogo Philippe Descola ha trazado un mapa de la combinatoria de opciones de que dispone el ser humano para

establecer relaciones con su entorno. En un momento de su análisis se centra en lo que denomina los “esquemas de la práctica”, que se encuentran a medio camino entre las impresiones sensibles particulares y los conceptos generales, haciendo por tanto de mediadores entre el agente y su entorno. Establece entonces un paralelismo entre la cognición humana y la técnica, para lo que distingue primero entre los “modelos clásicos” de programación y los “modelos conexionistas”, es decir, el paradigma en el que se basan las redes neuronales, el subtipo de la IA que desde hace poco más de una década está provocando la actual revolución tecnológica (Descola, 2005, pp. 186-187).

El “modelo clásico” funcionaba a partir de una lista de instrucciones que se suministraba de forma explícita a la máquina para que realizara los cálculos. En cambio el nuevo modelo “está constituido por un conjunto de redes electrónicas la interconexión de las cuales se establece selectivamente en función de la naturaleza e intensidad de los estímulos recibidos”. Por tanto, hay una oposición entre una forma de conocimiento que se basa en establecer “reglas explícitas” que formulen o capten alguna regularidad del ambiente y otra donde la máquina reacciona rápidamente a los cambios en el entorno y “parece obedecer a reglas formales sin que ninguna estipulación se haya introducido en el modelo”. Así pues, para Descola, la oposición dentro del ámbito técnico entre un “saber declarativo” (o proposicional) y un “saber-hacer” basado en el aprendizaje a partir de la práctica - entre el *know-what* y el *know-how* - refleja la misma oposición con la que se encuentra el etnólogo cuando quiere estudiar las reglas de comportamiento de una cierta comunidad, algunas de las cuales son explícitas pero otras en cambio están incorporadas de forma implícita (en forma de hábitos) en la práctica de los individuos.

La comparación que realiza Descola es pertinente por dos motivos. En primer lugar, porque una de las conclusiones que se

desprenden de ella es que el umbral alcanzado por la IA en el campo del conocimiento consiste en haber pasado de un saber declarativo a un saber procedimental, una capacidad hasta ahora exclusivamente humana que por primera vez ha sido exteriorizada con éxito en soportes técnicos. En segundo lugar, si el dominio del hábito (*know-how*) ha sido alcanzado por las máquinas ¿esto implica que toda una parte de la competencia humana ha de ser delegada en sus prótesis? En otras palabras, si el conocimiento basado en competencias (interiorización de hábitos prácticos, saber-hacer, etc.) ya no es una prerrogativa exclusivamente humana ¿quiere esto decir que antes o después seremos sustituidos en este ámbito y por tanto tendremos que emigrar a otro nicho cognitivo? ¿A cuál?

Edward Lee ha propuesto recientemente un paralelismo similar (2020, pp. 66-68). El progreso en la IA se ha basado en la introducción de la retroalimentación (*feedback*) en el mecanismo interno de los programas: la generación anterior se basaba en los “sistemas expertos”, donde los humanos codificaban primero y suministraban después a la máquinas las reglas explícitas que capturan el conocimiento de cómo se comporta un cierto ámbito; en la nueva generación, en cambio, “las representaciones de conocimiento se aprenden a partir de los datos”, no obedecen a órdenes sino que se extraen las regularidades a partir de muchos casos de ejemplo, tratando de ajustar progresivamente la respuesta efectiva del programa delante de los datos de entrada a un modelo esperado de cómo tiene que ser esa respuesta. De aquí el bucle de feedback entre los datos de entrada y salida (para una descripción en la misma línea de las diferencias entre el “paradigma simbólico” y el “conexionista” de la IA véase de Carolis, 2004, pp. 112-123).

Podríamos añadir por nuestra cuenta a esta idea la mitad que le falta para completar la analogía. El paso realizado por la IA desde la aplicación de un saber explícito a la producción de conocimiento implícito mediante la exposición a situaciones cambiantes ¿no calca

el paso dado por los humanos entre el *know-what* y el *know-how*, entre el modelo pedagógico de un conocimiento basado en contenidos (considerado hoy en día obsoleto) y un conocimiento basado en competencias, que ha funcionado (al menos hasta ahora) como modelo de todo conocimiento deseable? Precisamente, si el saber de contenidos viene considerándose obsoleto desde hace algunas décadas es muy posible que esto se deba a que ya había sido exteriorizado por las máquinas de una generación anterior. Como se suele decir ¿de qué sirve el aprendizaje memorístico en la época de Google y Wikipedia, cuando cualquier contenido está al alcance de cualquiera en cualquier momento? De la misma manera, debemos hoy preguntarnos *¿de qué sirve el conocimiento competencial en la época de la IA, cuando cualquier rendimiento competencial está ya (casi) al alcance de cualquiera en cualquier momento?*

Si por competencia entendemos “la capacidad de actuar en situaciones complejas e imprevisibles aplicando los conocimientos aprendidos” en lugar de “reproducir el conocimiento transmitido” (Sanmartí, 2020), entonces, una vez que la IA parece ser capaz de reproducir este tipo de conocimiento ¿deberíamos cambiar de modelo y buscar otro todavía no exteriorizado por la técnica? La IA ha realizado en poco tiempo el cambio que supuestamente tenían que realizar los humanos en el paso de una generación (la de la sociedad industrial) a la siguiente (sociedad postindustrial). Por tanto, las máquinas, que han pasado de un modelo de conocimiento “dirigido por las instrucciones” a uno “dirigido por los datos”, son hoy ya *competenciales*, es decir, capaces de actuar en un entorno postindustrial: cambiante, incierto, contextual, etc. Como sostiene Anthony Elliott “mientras que las máquinas automatizadas convencionales permanecían fijas en un lugar y eran programadas para tareas específicas y repetitivas, las nuevas tecnologías son móviles, sensibles a la situación y pueden adaptarse y comunicarse con su entorno” (Elliott, 2018, p. 21). Esto plantea el debate sobre si

la nueva oleada de automatización implicará la destrucción de puestos de trabajo humano (y no solo de baja cualificación profesional) o si la creación de nuevas profesiones compensará la destrucción de los antiguos puestos de trabajo (véase el equilibrado balance entre “escépticos” y “transformacionistas”, decantado finalmente hacia los primeros, en pp. 55-77). ¿Qué otro cambio o adaptación impone este nuevo entorno a la siguiente generación? Si las máquinas actuales ya son competenciales ¿qué nos tocará ser a nosotros? ¿Cómo interactuar con un entorno formado por máquinas que saben interactuar con su entorno? El reto por lo que respecta al conocimiento es entonces el siguiente: después de la transición del modelo memorístico al competencial ¿tendrá que seguir ahora una nueva transición? (y ¿hacia dónde?).

La respuesta es sí y no. Comencemos por el sí. Podríamos establecer un paralelismo entre la tecnología digital del siglo XX y la IA de principios del XXI. La digitalización del siglo pasado se basaba en la desmaterialización de los contenidos: texto, audio e imagen se convertían en información indiferenciada, una “gelatina de información” (como decía Marx respecto al trabajo), que permitía pasar cualquier contenido del mundo real al virtual. Este proceso provocó un aumento del ritmo de rotación de contenidos y consecuentemente dio lugar a formas de atención flotante y una memoria de trabajo más reducida. El modelo de las competencias surge como reacción a este tipo de entorno: una forma de adquirir conocimientos que sea compatible con formas de atención superficial y discontinua es precisamente la que tiene lugar a través del hábito o la ejercitación de capacidades (implícitamente, a través de la práctica, y no explícitamente, debiendo estar atento a lo que dice otro).

La IA, en cambio, hoy ya es capaz de hacer de forma automática una buena parte de aquello que hasta ahora se tenía que aprender a través del hábito, como, por ejemplo, diferentes formas de organizar

información (clasificar, relacionar, encontrar semejanzas, redactar, etc). Por tanto, en este nuevo entorno asistido por herramientas con IA, habrá que desarrollar aquellas capacidades que saquen provecho de estas nuevas herramientas y sepan hacer lo que ellas no saben hacer por sí mismas, sin orientación externa. Para concretar en qué podrían consistir estas *supercompetencias*, con una relación con las competencias como la que éstas tenían con los contenidos, podemos tomar como punto de partida la diferencia cualitativa entre las capacidades humanas y las técnicas. La IA, una vez ha logrado con éxito el *learning by doing*, necesitará todavía el juicio humano. La tecnología actual sabe hacer lo que antes requería una coordinación humana con las máquinas, por tanto, en la época de la automatización del conocimiento ahora la coordinación tendrá que cambiar para adaptarse a las nuevas prestaciones cognitivas del sistema de aparatos. Las capacidades humanas se tendrán que integrar y sincronizar con el sistema exteriorizado de capacidades, tal como lo hicieron en el pasado a través del modelo de aprendizaje por competencias.

Consideremos ahora el no: no hay nada más allá del marco de las competencias. Solo hay los contenidos y lo que hacemos con ellos. De la misma manera que el conocimiento competencial no ha desplazado del todo al conocimiento de contenidos sino que ha supuesto sólo un cambio de prioridades, dando lugar a un modelo híbrido, tal vez estemos ahora en la misma situación. Más que ser sustituido, el modelo de las competencias tendrá que ser reconsiderado.

Conocimiento encarnado

Una tal reconsideración podría pasar por revalorizar lo que se conoce como “conocimiento encarnado” (Lee, 2020, p. 105). La capacidad

de conocer no reside únicamente en el cerebro sino también en el cuerpo. En el ser humano el conocimiento no es separable del instrumento con el que se adquiere, el propio cuerpo, a diferencia de la IA, en la que un mismo software puede ejecutarse en diferentes hardware, lo que se conoce como “realizabilidad múltiple” (Nowotny, 2021, pp. 75-78). Esta es, por tanto, una primera diferencia que singulariza al ser humano delante de la IA. ¿Por qué es importante? Es cierto que las tareas que aprendemos a hacer, a través de la interacción entre el cuerpo y el cerebro, las pueden hacer hoy también algunas máquinas pero es importante mantener este aprendizaje ya que es la condición para aprender a hacer otras tareas que (al menos de momento) no pueden hacer las máquinas. Por tanto, sustituir aquellas tareas por su versión mecánica sería un error ya que interrumpiría un proceso de aprendizaje a más largo plazo. Por ejemplo, el cálculo mental hace tiempo que puede ser delegado en las máquinas pero todavía se enseña porque ayuda a aumentar la memoria de trabajo, lo que es útil posteriormente para realizar otras tareas más complejas.

Richard Sennett, un gran defensor de este tipo de conocimiento, distingue a este respecto entre un mal y un buen uso de la tecnología (Sennett, 2009, pp. 55-62). El mal uso es el que impide el aprendizaje por repetición, basado en hacer y luego observar lo que se ha hecho, ya que interrumpe el bucle acción-observación que es una de las fuentes elementales de la experiencia humana. Las máquinas “separan la cabeza de la mano” - la comprensión mental, puramente abstracta, del aprendizaje manual, realizado en primera persona - y rompen así el bucle de retroalimentación entre hacer y pensar, ya que suministran el resultado separado del proceso del que surge, lo que disminuye el grado de comprensión. La máquina se convierte en un proveedor de datos correctos para la comprensión mental, pero al precio de cortocircuitar el bucle a través del cual validamos un dato como correcto o incorrecto. Como dice Sennett, nos volvemos

testigos pasivos de una competencia que va en aumento pero en la que no participamos. Por tanto, la máquina sustituye el proceso de adquisición de conocimiento y acaba impidiendo la propia intervención y exploración de un campo de posibilidades, lo que dificulta la autoinstrucción y un conocimiento activo y consciente.

La misma oposición que describe Sennett -entre una forma de conocimiento que depende de un cuerpo que percibe y actúa y otra donde hay un aislamiento relativo entre la obtención de datos (*input*) y el procesamiento de información (*output*)- es la que describe Descola (2005, pp. 325-329) cuando opone la “cognición ecológica” al modelo de la “máquina de Turing” (es decir, el ordenador). Los ordenadores, incluso los que son “capaces de modular sus acciones siguiendo un aprendizaje autónomo” (es decir, la IA), se basan en “modelos conexionistas” de procesamiento de información que presuponen una ontología cartesiana: la separación entre la unidad de control y los subsistemas de sensores y actuadores refleja así la separación entre espíritu y cuerpo. De un lado la obtención de datos, del otro el procesamiento o la capacidad de cálculo como un nuevo Ego cartesiano.

En cambio, una forma alternativa de pensar el conocimiento es a través de la “cognición incorporada o ecológica”. El conocimiento humano “es función de un cuerpo que percibe y que actúa”. Conocer no es meramente obtener información del ambiente, de una forma neutral o pseudo-objetiva, sino percibir estímulos del ambiente que puedan traducirse en una acción posible. Las preguntas que le hacemos al entorno están condicionadas por las posibles respuestas que le puede suministrar nuestro cuerpo. Parafraseando a Marx, cada organismo sólo se plantea los problemas que es capaz de resolver. El conocimiento humano está siempre situado, es decir, sesgado por sus propios instrumentos cognitivos. Por tanto, depende de dos factores: de un lado, de las capacidades de percibir y de actuar, es decir, de las posibilidades limitadas que admite el aparato perceptivo y locomotor

del cuerpo humano; de otro lado, del vínculo o libre juego que se establece entre estos dos aparatos, que no son como los subsistemas abstractos y, por tanto, independientes de una máquina (cada uno con una definición autónoma de la tarea que ha de cumplir) sino que están definidos uno en relación al otro. La percepción no tiene como objetivo obtener datos y la unidad de control producir respuestas sino que la percepción trabaja en vistas a una acción posible y la acción en vistas a la percepción. Esta retroactividad entre los dos aparatos hace de ellos una unidad donde cada polo es a la vez condición y resultado del otro: uno proporcionando datos para una acción posible y el otro produciendo acciones con un resultado que a su vez se constituye en un nuevo dato observable, cerrando el bucle que define la acción humana.

Señal y ruido

Que el mundo no está formado por propiedades objetivas (datos, información) sino que viene filtrado o interpretado por las posibilidades de acción que se le presentan al individuo es también uno de los temas clásicos de reflexión de la antropología filosófica europea del siglo XX, a partir primero de Jakob von Uexküll y después de Arnold Gehlen o Helmuth Plessner, autores que intentaban desmarcarse del intento más tradicional de dar una definición de una esencia humana (ahistórica y no situada en un contexto práctico) y que vieron en el giro hacia la acción la forma de desarrollar un estudio del ser humano menos cargado de presupuestos metafísicos y entidades inobservables (Rehberg, 1983, pp. 385-402).

Para Uexküll, por ejemplo, el sentido que tienen los estímulos y cómo se percibe el mundo depende del catálogo más o menos limitado de acciones que puede realizar cada organismo. Es el “tono

efectual” (Uexküll, 2016, p. 105, p. 144) de una percepción lo que convierte a los estímulos en puntos de referencia para la orientación práctica en el mundo. Aunque diversas especies compartan un mismo espacio en realidad viven en diferentes ambientes, que recortan a partir de la diferente dotación instintiva y las peculiaridades de su sistema sensomotor. El “mundo” objetivo (*Welt*) se difracta entonces en una diversidad de “ambientes” (*Umwelt*), según las cosas se carguen, como signos, con significados que dependen de lo que se puede hacer con ellas.

La distinción entre mundo y ambiente ha sufrido sin embargo una cierta fluidez terminológica. Si para Uexküll no es todavía una oposición tajante sino una diferencia de grado que permite sostener una cierta continuidad entre el ser humano y el animal (capaces, cada uno a su manera, de recortar mediante la acción un ambiente a partir del mundo ilimitado), en cambio para Gehlen servirá para marcar una diferencia sin matices entre el mundo humano y el ambiente animal (apertura contra clausura), aunque en su obra posterior esta oposición se irá matizando progresivamente. Finalmente, Plessner se muestra más ambiguo cuando sostiene que lo propio del hombre es hacerse “un ambiente a partir del mundo”, donde mundo y ambiente parecen haber intercambiado su significación respecto a la intención inicial de Uexküll. En todo caso, a pesar de las diferencias, lo que estos autores tienen en común es la afirmación de una paradoja antropológica irresoluble (de Carolis, 2017, pp. 55-61): entre *apertura* (la exploración de nuevas posibilidades propia del “mundo”) y *protección* (el límite y la negación de la contingencia propia del “ambiente”).

La importancia actual de la distinción entre ambiente y mundo va más allá del ámbito especializado del que procede y puede captarse mejor si se la interpreta como “una diferencia entre dos modalidades de elaboración de la información” (de Carolis, 2004, pp. 56-61). Más concretamente, entre dos formas de relación entre señal y ruido.

Manejarse en un ambiente significa que la distinción entre lo que tiene significado (señal) y lo que no lo tiene (ruido) es relativamente fija y presenta poco margen de variación. La separación entre la figura y el fondo informe depende de una dotación biológica (genética, instintiva, sensomotora, etc.) que admite pocas variaciones. Desempeñarse en un mundo en cambio quiere decir que lo que vale como señal (y por tanto significa algo) no está dado de una vez por todas sino que hay un límite móvil que separa pero a la vez comunica señal y ruido, lo que es relevante o significativo de lo que no es. Aquí la selección de lo que vale como señal puede volver a producirse de nuevo a partir del ruido, dado que este funciona como una reserva potencial de significado a la espera de ser activada y transformar el ruido en un nuevo orden.

La antropología filosófica europea ha insistido de forma preponderante sobre el polo de apertura al mundo como rasgo dominante de la condición humana pero, como señala de Carolis, desde el principio de su reflexión supo ver en ella un bucle o reenvío continuo entre el polo de la apertura y el polo del repliegue: recortar un ambiente estable y reabrir de nuevo el acceso a un mundo, poner un límite ante el exceso de sensaciones y desplazar ese corte para rehacerlo en otro lugar. Ambos coexisten en la criatura humana y hacen de su condición un campo de tensiones productivas e irresolubles.

¿Y la inteligencia artificial? ¿Tiene un ambiente o un mundo? ¿Cuál es la relación que establece entre señal y ruido? ¿Hay un límite fijo o una cierta indistinción entre ambas? ¿Muestra algún grado de apertura al mundo o el entorno con el que interactúa se reduce a un conjunto de datos objetivamente definido?

Para empezar hay que recordar que una red neuronal no elige los datos con los que se entrena sino que es el programador quien de forma externa hace una selección y se los proporciona. Por tanto, la delimitación de qué vale como dato significativo para el aprendizaje

de la red y de cuál es su significado (qué valor de salida le corresponde) está fijado de antemano. No está fijado, ciertamente, el valor o peso que un dato de entrada adquiere dentro del proceso de aprendizaje para producir una salida, y esa flexibilidad es la que permite precisamente el aprendizaje. Se podría decir que la máquina no necesita criterio de selección, el cual es un índice de la debilidad humana, ya que para el ser humano todo es potencialmente significativo y para orientarse necesita hacer un corte, aunque sea provisional y revisable, para reducir la complejidad de su entorno y ser operativo. La máquina en cambio puede calcular a partir de todos los datos disponibles. Sin embargo, en su caso es el objetivo el que funciona como criterio de selección de los datos de entrada (aquellos que sirvan para aprender a realizar mejor una cierta tarea), un objetivo que le viene marcado desde fuera (como veremos en la sección siguiente).

¿Qué es aprender entonces para el ser humano y para la IA? Para el primero, desplazar el límite entre señal y ruido, de forma que lo no significativo se vuelva significativo y donde solo había ruido se perciba una nueva forma de orden. Para la segunda, aprender no tiene que ver con captar un sentido sino con aprender a producir un resultado correcto. Para ello la IA se sirve de la agregación de unidades de cálculo subsimbólico (las neuronas de una red, cada una de ellas ciega al significado), de las cuales el significado emerge como el resultado final, de acuerdo con parámetros fijados de antemano que son revisables para una apreciación humana externa pero no para la propia máquina. Por tanto, en ambos casos el aprendizaje discurre en la misma dirección pero en sentidos opuestos (de Carolis, 2004, p. 60): para la máquina aprender significa alcanzar una mayor concretización en su respuesta al entorno (fijación de la distinción señal/ruido, es decir, reducción del ruido al mínimo), para el hombre aprender implica lograr una mayor ductilidad (exponerse al ruido para transformarlo en señal).

La IA no tiene entonces capacidad de reabrir la distinción entre señal y ruido y ese es precisamente su punto fuerte, siendo su ceguera al significado de los datos el precio que tiene pagar para resultar eficaz. De forma más general podríamos decir que la respuesta de un organismo será tanto más eficaz cuando más cerrado esté al ruido inútil (ambigüedad entre señal y ruido). Inversamente, un organismo exhibirá tanta más potencia interpretativa cuanto menos presión tenga para responder de forma eficaz. Lo cual parece conducir a una forma de complementariedad entre la inteligencia natural y la artificial -entre interpretar y calcular, mundo y ambiente, apertura y protección- que no solo no deshace sino que refuerza la paradoja antropológica.

De hecho, se podría decir, la técnica es la exteriorización de uno de los polos de la paradoja interna que constituye al ser humano, de ahí que siempre haya servido de espejo de reconocimiento. El bucle entre significación y eficiencia, del cual uno de los elementos se incorpora al mundo externo en forma de dispositivo técnico, amplía su alcance excediendo la esfera de la subjetividad para incluir lo que está más allá de ella. La técnica aparece así para el ser humano como una forma de relación consigo mismo a través de lo otro de sí. Esto implica que el intento de marcar una diferencia demasiado tajante entre el ser humano y la técnica está siempre desencaminado ya que la diferencia remite al interior de la propia condición humana. Este movimiento de ida y vuelta entre el ser humano y la técnica se puede reconocer de nuevo en el bucle entre práctica y producción.

Producción y práctica

En la *Ética a Nicómaco* Aristóteles distinguía entre dos tipos de virtudes o facultades humanas: las naturales, que preceden su ejercicio (como ver u oír), y las morales (o que pertenecen al

carácter), que son resultado de la propia acción del agente. “Los hombres se hacen constructores construyendo y citaristas tocando la cítara. Pues bien, de esta manera nos hacemos justos realizando acciones justas y valientes” (1103a). Posteriormente, cuando trata las facultades del intelecto, hace una nueva distinción entre dos subtipos: la *poiesis* (facultad propia de la esfera de la técnica) y la *praxis* (excelencia humana en general, sin delimitación de ámbito). Esta distinción entre dos formas de conducta humana -la producción y la práctica- puede resultar útil, después de un salto de más de dos mil años, para describir la diferencia actual entre las competencias humanas y las artificiales. ¿Se podría decir que la IA funciona según el modelo de la producción y la acción humana según el de la práctica?

Tal como sostiene Aristóteles, “ni la realización [*praxis*] es fabricación [*poiesis*] ni la fabricación realización” (1140a). La *producción* (o fabricación) coincide, a grandes rasgos, con la acción instrumental, mero medio para un fin (la realización de un producto) al que está subordinada y del que se puede separar de forma clara. La acción es lineal y su objetivo es el resultado. En la *práctica* (o realización), en cambio, el fin es la actividad misma, el hecho de realizarla bien o de forma excelente.

El perfeccionamiento técnico de la actividad productiva en el último siglo y medio, con la creciente tecnificación de diferentes ámbitos de acción (el trabajo, la guerra o el arte), se basa efectivamente en el momento de la producción: importa primeramente la *eficacia* con que se logra un resultado, prescindiendo de cualquier otra consideración. En cambio la práctica crea su propio sentido, que no es previo ni exterior sino que lo segrega, por así decir, la acción durante su ejercicio.

Desde el punto de vista de la práctica, la acción no es solo un medio para conseguir un objetivo sino un medio que, al mismo tiempo que persigue un fin, ayuda a revelar en qué consiste ese

mismo fin que se busca. Cada acción, en su particularidad específica, una vez realizada ejemplifica y, por tanto, da indicaciones de cómo sería un modelo logrado de esa acción. El fin, por tanto, es inseparable del medio, es en cierto sentido producido por el propio medio, que le sirve de guía. De nuevo, podría decirse también ahora con Marx que cada agente sólo se propone aquello que es capaz de realizar. El fin establece los medios necesarios para su consecución pero los medios redefinen el sentido que hay que darle al fin, dan acceso a nuevos aspectos insospechados suyos, por lo que modifican nuestra concepción de qué significa aquello que buscamos. Este acaba siendo el sentido de la acción: el conocimiento que la acción procura al agente sobre lo que busca al actuar (en el límite, qué quiere decir para el agente hacerse mejor, lo que los griegos llamaban “virtud”).

La acción presenta entonces la conocida forma de un bucle entre medio y fin: uno quiere actuar bien para ser virtuoso y quiere ser virtuoso para actuar bien. El valor de la actividad no se mide solo por sus efectos (*qué*) sino también por la modalidad interna de la propia acción (*cómo*). La acción es un medio para un fin, pero a la vez la acción como realización de un fin es usada como medio para saber en qué consiste ese fin. El propio medio es instrumentalizado en una acción de segundo grado con un rendimiento epistémico (procura conocimiento) y no solo práctico (realiza objetivos).

De lo anterior se sigue entonces que en la práctica no hay una definición previa del objetivo de la acción (un modelo óptimo) sino que la definición surge y se enriquece progresivamente por medio de la acción. Por tanto, la técnica trata la acción como producción y está dirigida a la *optimización* del producto. En cambio el ser humano puede preocuparse del sentido de la acción, que está relacionado con un proceso de *autoperfeccionamiento* o mejora de la propia calidad humana (para una interpretación de la diferencia técnico-humano a partir del par *poiesis-praxis* y sus variantes véase de Carolis, 2004,

pp. 38-44 y Bodei, 2022, pp. 236 y ss.). A través de la acción, logra resultados; a través del conocimiento que le brinda la acción se pone de camino hacia su propia humanidad.

¿Qué tiene que ver todo esto con la IA? Podríamos decir que la IA es platónica mientras que el ser humano es aristotélico. El algoritmo con el que funcionan las redes neuronales actuales requiere de la definición previa de un óptimo de su rendimiento: es necesario saber cuál es la respuesta esperada para poder medir la distancia entre la respuesta que produce la máquina y la que tendría que producir. De manera que el objetivo del algoritmo de aprendizaje es maximizar el éxito en las respuestas que genera (o, lo que es lo mismo, minimizar el error). Por tanto, la red neuronal puede optimizar su rendimiento solo porque parte de una definición previa de lo que es un óptimo (que es definido por el programador cuando le suministra el conjunto de datos de entrenamiento). La IA requiere de estos tres ingredientes: la definición humana de un óptimo, una métrica que indica si se acerca o no a su rendimiento óptimo (lo que se conoce como “función de pérdida”) y la modificación interna de los pesos de la red para minimizar esa métrica. A pesar de la gran variedad de técnicas que utiliza la IA actual, todas se basan en el “modelo estándar de la IA”, es decir, “un algoritmo que optimiza un objetivo prefijado” (Russell, 2019, pp. 60-64). El modelo alternativo, el de la “IA de propósito general”, no está aún a la vista (sobre la insistencia de una “IA benéfica” bajo el control de objetivos humanos véase también Tegmark, 2018, pp. 61-63).

Esta actividad encaja con el modelo de la producción (la acción como un medio para conseguir una finalidad definida previamente). Por tanto, tendrá éxito allí donde sea posible definir un óptimo pero fracasará donde lo no sea, no solo porque aún no sabemos cómo hacerlo sino porque sencillamente no es posible hacerlo, porque es consustancial a la acción humana el hecho de no tener óptimo (una definición de una vez por todas). Y no tiene óptimo precisamente lo

que tiene sentido, los propósitos que extraen su definición de forma progresiva únicamente a través de su propio ejercicio: la práctica.

Procesamiento de información

Como en el caso de la memoria, el procesamiento de la información se encuentra con una alternativa similar. La IA es competente encontrando relaciones dentro de un alto volumen de información aunque sea heterogénea y no estructurada. Funciona bien con un alto nivel de ruido. El ser humano, en cambio, encuentra relaciones cuando el volumen es relativamente bajo y la información está altamente estructurada (Lee, 2020, p. 160).

Esto añade un nuevo nivel de complementariedad, por ejemplo, en el uso de asistentes personales (búsqueda, recomendación, etc). Cada vez estamos más habituados a realizar nuestras tareas diarias siendo asistidos por algoritmos que desbrozan la maleza digital (un terreno con un alto nivel de ruido, tanto por el volumen como por la heterogeneidad) y pueden facilitar así en un segundo momento el análisis y la decisión humana. La situación es paradójica: la digitalización, que ha llenado nuestro entorno con un alud de contenidos de una altísima velocidad de rotación, ofrece a la vez las herramientas para orientarnos en este caos. Por una parte, ha roto el bucle entre acción y observación que es propio de la experiencia humana. Gehlen ya había detectado en los años cincuenta del siglo pasado la “ruptura del círculo de la acción” debido al desarrollo tecnológico, de forma que la percepción del entorno ya no es capaz de guiar la acción y los resultados de la acción a su vez son incapaces de dar lugar a nuevas percepciones (Gehlen, 2004, pp. 42-51 y Postman, 1991, pp. 73-75). Pero, por otra parte, la nueva generación de asistentes con IA ha ayudado a cerrar de nuevo el bucle ayudándonos a obtener sentido a partir del ruido creciente que

caracteriza a nuestros entornos digitalizados.

Esta situación de complementariedad, sin embargo, es ambigua y también admite una lectura más escéptica ya que en cierto sentido hemos empezado ahora a ser sustituidos por nuestros asistentes en un rasgo esencial de nuestra relación con el entorno: no ya en esta o aquella capacidad sino ahora en la capacidad de segundo orden de adquirir experiencia mediante bucles de retroalimentación (*feedback*) entre percepción y acción. Ya no seríamos capaces de tomar decisiones o juzgar sobre la verdad de una región del mundo sin la mediación del sistema de aparatos que percibe el mundo por nosotros (filtra, selecciona, recomienda) y, en definitiva, guía nuestra acción. Este es el núcleo de la crítica de Èric Sadin, que advierte del nuevo régimen de verdad que instaura la nueva generación de asistentes conducidos por IA (Sadin, 2020, pp. 65-70). Esta interpretación deplora la pérdida de control humano sin advertir que la exteriorización de una capacidad siempre implica en cierta medida una transferencia de control humano a los aparatos, que funcionan según sus propias reglas, y, por tanto, una cierta dependencia. El problema con la tesis del control sigue siendo el presupuesto subyacente de un sujeto que controla y, por tanto, se sitúa al margen de los cambios que no sean iniciados por él mismo. Este presupuesto debe ser cuestionado. Parafraseando a Walter Benjamin en *Calle de dirección única* (2010, p. 88), no se trata ya de mantener el control de la técnica sino de controlar la relación entre el ser humano y la técnica. Esto quiere decir que de lo que se trata es de buscar una adaptación del ser humano a los aparatos que no implique subordinación sino que sirva de medio a la realización de las mejores posibilidades de la técnica actual (no previstas por él pero inscritas en los propios dispositivos y disponibles para su uso). Explorar el nuevo papel que le corresponde como sujeto dadas las actuales condiciones técnicas de forma que pueda sacar el máximo partido de estas condiciones para sus propios fines (volveremos

sobre este punto en la sección final sobre coevolución).

Un buen ejemplo de esta relación de complementariedad humano-máquina lo ofrece la cuestión de la “explicabilidad”. La IA actual se comporta como una “caja negra”: es capaz de ofrecer resultados útiles para asistir decisiones humanas pero no es capaz de explicar la validez de esos resultados (Coeckelbergh, 2022, pp. 100-106; Sadin, 2020, pp. 76 ss). Por tanto, aquí el riesgo es que las verdades que ofrece deban tomarse como hechos sin justificación. Esto a su vez genera otro problema, dado que la autoridad que tendemos a concederles a las máquinas tiende a invisibilizar el hecho de que han sido diseñadas por humanos con objetivos humanos y para lograr beneficios para como mínimo algunos humanos. Esto implica una ilusión opuesta a la del control, es decir, la de decisiones que no están bajo control humano, muy conveniente para blindar las decisiones contra posibles quejas y reclamaciones. Es lo que Postman había criticado como “transferencia de responsabilidad” en la década de los noventa en referencia a la aparición del ordenador, al que tendemos a atribuir rasgos antropomórficos (2018, pp. 154-158).

En todo caso es cierto que en la IA el resultado final surge de la agregación de una multiplicidad de cálculos distribuidos en las unidades o “neuronas” que forman la red pero no hay ningún sitio de la red neuronal al que dirigirse en búsqueda de una explicación. Pero esto implica que no puede dejarse al arbitrio de la máquina la decisión última (sobre un diagnóstico médico, la concesión de un crédito o la prima de un seguro) y compete al usuario humano aportar la justificación y la responsabilidad última del uso de la IA.

La actual generación de máquinas está introduciendo un índice de opacidad en la superficie social, donde su uso va en aumento, que hace que cada vez sea menos realista pensar en la posibilidad del control humano (Dyson, 2020, pp. 246-253). La ley de protección de datos a nivel europeo, por ejemplo, sostiene que todos los

ciudadanos tienen derecho a obtener una explicación de las decisiones que les afecten y estén basadas en procesos automáticos (Lee, 2021, p. 99), pero en vista del hermetismo intrínseco de esos procesos la medida es mayormente papel mojado. Sin embargo, aquí de nuevo debe sustituirse la perspectiva del control por la de un reparto equitativo de roles (obtención de resultados, por un lado, y justificación y evaluación de consecuencias, por otro) que permita que ambos agentes trabajen juntos.

Algo parecido pasa con el tema del procesamiento de información, donde hay que distinguir (Lee, 2020, cap. 11) entre *correlaciones* y *relaciones*. La IA es muy eficaz con un volumen muy elevado de datos pero solo indica correlaciones (es decir, coincidencias estadísticas) que no necesariamente corresponden a relaciones (un vínculo causal, por ejemplo). En todo caso, detectar correlaciones puede ser útil si después hay una validación humana que verifique si detrás hay una auténtica relación o se trata de una mera coincidencia. De nuevo aparece aquí la complementariedad.

Por último, se puede indicar otra diferencia en la relación que se establece en cada caso entre los datos y la conclusión. La IA funciona según el modelo del “conocimiento guiado por los datos” (*data-driven knowledge*), que consiste en aprender desde cero a partir de datos (sin la mediación de conceptos). Esto provoca que en algunas tareas su rendimiento sea todavía muy bajo, precisamente allí donde el conocimiento requiere ser encapsulado en diferentes niveles de sentido (conceptos). Esta forma de trabajar de la IA contrasta con la manera en que lo hace la inteligencia humana, que extrae predicciones a partir de información insuficiente (en los ámbitos, por ejemplo, donde no hay la posibilidad de obtener datos), es decir, añade alguna cosa más de lo que hay en los datos de entrada. Por ejemplo, a partir de la coherencia con otros conocimientos previos (como el conocimiento tácito sobre el mundo adquirido por experiencia directa). El conocimiento humano (Russell, 2019, p. 84)

es estratificado (distingue entre capas o niveles de organización abstractos) y acumulativo (cada capa permite acceder a nuevos niveles de abstracción en un proceso progresivo).

Conclusión: el futuro exige sabiduría

Hemos visto algunas diferencias entre el ser humano y la IA que apuntan a explorar una forma de relación complementaria, donde cada parte compensa sus puntos débiles con las fortalezas de la otra. Los rasgos diferenciales de cada tipo de inteligencia no excluyen sino que requieren de su polo opuesto. Esta forma de complementariedad humano-máquina se acerca a lo que en biología evolutiva se conoce como “coevolución”. Aunque existen diferentes formulaciones del concepto -ya sea que indique la relación entre dos especies, entre una especie y su entorno o entre dimensiones distintas dentro de la misma especie (biología y cultura)¹- en todo caso, el núcleo común de la idea señala la dependencia mutua entre ámbitos en principio heterogéneos. Por lo que respecta al proceso de hominización, apunta al bucle de determinación recíproca entre la biología (la anatomía, el sustrato material) y la cultura (la conducta, las posibilidades de hacer o pensar).

Contra el tópico transhumanista (Tegmark, 2018, pp. 41-45) de que la evolución natural queda superada cuando el ser humano alcanza un cierto grado de desarrollo cultural o tecnológico (como si la invención cultural permitiera quedar al margen de las presiones selectivas del medio natural), la coevolución defiende en cambio una

¹ Sobre la aplicación de la idea al caso particular de la hominización véanse, por ejemplo, la teoría de la “superposición” entre biología y cultura (Geertz, 2011, pp. 51-54 y 66-71); la “coevolución gen-cultura” (Wilson, 2012) o el condicionamiento de la transmisión cultural sobre la evolución genética de la especie humana (Boyd, 2018). Debo agradecer a Josep Maria Lloró por darme a conocer estas referencias.

interacción continua entre naturaleza y cultura. No hay una línea temporal de desarrollo sino un bucle.

¿Cuál es la virtualidad crítica de esta idea? En primer lugar, la de acercarse sin simplificar a una realidad compleja, donde las relaciones no se basan en la superación o progreso entre un nivel inferior y uno superior sino en la tensión entre opuestos. En segundo lugar, ayuda a poner distancia ante la idea de una autonomía prometeica del ser humano, capaz de superar todos los límites y condiciones. Idea que segrega una variada tipología de fantasías de control (lo que incluye también a su opuesto, las fantasías distópicas de la amenaza ante la falta de control). Desde el punto de vista de la coevolución el ser humano aparece como un ser situado o sujeto a condiciones: lo que hacemos nos hace. Lo que inventamos para superar las viejas condiciones solo nos libera al precio de sujetarnos a otras nuevas resultado de nuestra acción. Condición y resultado forman un bucle irresoluble.

El ser humano es resultado de sus propios productos o, si se quiere, producto de sí mismo a través de su acción sobre el mundo externo. Es lo que sostiene Michel Serres (2014, pp. 62-68; p. 286) cuando describe el proceso de “autohominización” formado por dos polos: el “exodarwinismo” (la exteriorización adaptativa de capacidades) y la “retroacción” del efecto de los objetos producidos sobre nuestras capacidades. O también Bernard Stiegler (2016, pp. 84-87; p. 102) cuando habla de la “exosomatización” formada por dos momentos de acción recíproca entre exteriorización (técnica) e interiorización (psíquica y colectiva).

La idea supone además una ventaja práctica: contra el optimismo transhumanista (o su contracara, el miedo ante la pérdida de control) apela a la responsabilidad del momento presente ya que pone el acento sobre el siguiente paso a dar en el proceso coevolutivo: dadas las actuales condiciones técnicas ¿cuál es el resultado deseable y a la vez compatible con ellas? Ofrece orientación a un ser humano

constituido en buena medida a partir de la tecnología anterior sobre cómo debe actuar para aprovechar las opciones que le ofrece la más reciente: no tratando de competir con ella (la optimización, que corta el nudo que une los dos polos que forman la paradoja antropológica, reduciendo el ser humano a una parte de sí mismo) sino mediante la complementariedad entre capacidades y funciones heterogéneas (que conserva la tensión entre los dos polos de la paradoja que nos constituye). Es lo que sostiene Edward Lee cuando contrapone la hipótesis coevolutiva, que ha de permitir potenciar la propia humanidad a través del uso de la IA, al “creacionismo digital” (2020, p. 311), que supone una relación jerárquica, ya sea para aumentar nuestro control sobre la IA o para limitar el suyo (a través de una legislación más restrictiva o de manifiestos en favor de una IA ética, por ejemplo).

Para finalizar, vamos a intentar señalar qué aplicación a la educación puede tener la tesis de una coevolución entre humanidad y técnica. En primer lugar, una evidente a la luz de lo anterior: el ser humano debe especializarse en su propia humanidad. Esto significa, en primer lugar, ejercitar y reforzar las *supercompetencias* humanas y aprender a utilizar las competencias técnicas para resolver problemas de forma conjunta. Es decir: aprender a combinar el consumo de información disponible con la producción de conocimiento nuevo en la interacción entre la cabeza y la mano (rediseño, contextos de prueba y error); aprender a explorar el sentido de un problema o de una pregunta, no solo a producir soluciones correctas; aprender a utilizar los asistentes como una herramienta heurística que requiere después un trabajo de comprensión y síntesis insustituible; aprender a dar cuenta de lo que hacemos (por qué), no solo a seguir procedimientos (cómo) o producir resultados (qué).

En segundo lugar, si es cierto que “las habilidades y los contenidos” han quedado obsoletos en el ámbito de la educación,

dado que las máquinas los gestionan mejor que los humanos (Lee, 2020, p. 309), entonces ¿cuál ha de ser el objetivo de la educación? Una posible respuesta es que cada vez será más importante el estudio de la tecnología como forma de entender la sociedad (Lee, 2020, p. 310): se tienen que estudiar tendencias y no hechos, que pueden cambiar de un día para otro. Por ejemplo, las fuerzas ocultas que amenazan la privacidad, la divulgación viral de ideas o una historia reciente del mundo digital que evite la aceptación acrítica de la tecnología como un hecho incuestionable.

Se podría añadir la necesidad de un humanismo digital. El mundo se llena cada vez más de fenómenos que toman la apariencia de hechos, cerrados sobre sí mismos y aislados unos de otros, lo que impide captar su interacción y, por tanto, una comprensión adecuada e integral. El estudio de las humanidades, en un sentido generalista, puede ser hoy el mejor asistente de la humanidad en su relación con un mundo cada vez más extraño o ajeno. Los hechos que se presentan como independientes de nuestra intervención pueden adquirir un nuevo sentido cuando se conectan con otros, pasados o presentes, dibujando así un marco más amplio donde el mundo recupera su inteligibilidad y nosotros nuestra capacidad de intervención (ya que entendemos que el mundo no es independiente de nuestra acción). Así, Postman defendía una educación centrada en la coherencia para evitar el consumo pasivo de verdades como hechos, para lo que recomendaba poner énfasis en la historia. Los contenidos de cualquier asignatura debían incorporar un tratamiento de la historia de la propia disciplina (2018, pp. 247-257).

Por tanto, no que la técnica sea el asistente entre el ser humano y la naturaleza sino que la cultura (sin especificación de ámbitos, en lo que todos ellos tienen en común de reserva de significado y, por tanto, de potencial de interpretación de nuevos hechos) se convierta ahora en el asistente entre el ser humano y un medio técnicamente configurado. Solo así, a través de esta operación de segundo grado

(mediación cultural de la mediación técnica), podremos tal vez volver a cerrar, no al nivel del individuo sino del colectivo (y, en el límite, de la especie), el bucle entre actuar y comprender, fracturado debido a la disrupción tecnológica.

El futuro exige sabiduría, como sostiene Helga Nowotny (2021, pp. 126-133). Los retos actuales requieren mirar a la vez hacia delante y hacia detrás, vincular el futuro y el pasado, atender a la urgencia de los problemas a corto plazo y mantener una visión general, lo que incluye *saber releer el patrimonio cultural* a la luz de los problemas presentes y viceversa, en una iluminación recíproca.

La sabiduría es saber escoger. Hay que distinguir lo que es verdad de lo que no lo es y también lo que es eficaz de lo que no lo es, pero como mínimo igual de importante es saber apreciar lo que tiene valor y rechazar lo que no lo tiene. En un contexto de alta volatilidad de contenidos se vuelve difícil la previsión del futuro, incluso a muy corto plazo, por lo que ante la incapacidad de saber qué nos será útil más adelante se impone la tentación de conservarlo todo. Cuando el futuro falla como criterio de selección del pasado, como actualmente, se vuelve urgente *saber buscar otros criterios de selección* que nos ayuden a distinguir lo que es necesario de lo que es superfluo. Tener sabiduría, por último, también significa resistirse a la tentación del control y mostrar una cierta apertura para aprender de la nueva tecnología. Contra el miedo a la sustitución humana por la IA y la competencia por la primacía como única opción, hay que *saber complementar mutuamente las carencias y las fortalezas del ser humano y la tecnología*. Dice mucho de la dificultad y el volumen de los retos a los que tenemos que hacer frente actualmente que el objetivo de la educación tenga que ser, ni más ni menos, que el de *enseñar la sabiduría*.

Referencias

- Aristóteles (2001). *Ética a Nicómaco*. Alianza.
- Benjamin, W. (2005). *Libro de los Pasajes*. Akal.
- Benjamin, W. (2010). *Obras - Libro IV/vol. I*. Abada.
- Bodei, R. (2022). *Dominio y sometimiento. Esclavos, animales, máquinas, inteligencia artificial*. Alianza.
- Boyd, R. (2018). *Un animal diferente. Cómo la cultura transformó nuestra especie*. Oberon.
- Coeckelbergh, M. (2022). *Ética de la inteligencia artificial*. Cátedra.
- De Carolis, M. (2004). *La vita nell'epoca della sua riproducibilità tecnica*. Bollati Boringhieri.
- De Carolis, M. (2017). *La paradoja antropológica. Nichos, micromundos, disociación psíquica*. Quadrata.
- Delgado, J. L. (2016). *El origen del presente. La filosofía de la historia de Walter Benjamin*. Sequitur.
- Descola, P. (2005). *Par-delà nature et culture*. Gallimard.
- Dyson, G. (2020). *Analogia. The entangled destinies of nature, human beings and machines*. Penguin.
- Elliott, A. (2018). *The culture of AI. Everyday life and the digital revolution*. Routledge.
- Geertz, C. (2011). *La interpretación de las culturas*. Gedisa.
- Gehlen, A. (2004). *Die Seele im technischen Zeitalter. Gesamtausgabe. Band 6*. Klostermann.
- Lee, E. A. (2020). *The coevolution. The entwined futures of humans and machines*. MIT Press.
- Lipovetsky, G. y Serroy, J. (2015). *La estetización del mundo*. Anagrama.
- Nowotny, H. (2021). *La fe en la inteligencia artificial. Los algoritmos predictivos y el futuro de la humanidad*. Galaxia Gutenberg.

- Ong, W. J. (2016). *Oralidad y escritura: tecnologías de la palabra*. FCE.
- Postman, N. (1991). *Divertirse hasta morir. El discurso público en la era del "show business"*. Tempestad.
- Postman, N. (2018). *Tecnópolis. La rendición de la cultura a la tecnología*. El Salmón.
- Rappaport, R. A. (2016). *Ritual y religión en la formación de la humanidad*. Akal.
- Rehberg, K.S. (1983). Nachwort des Herausgebers. En A. Gehlen, *Gesamtausgabe (Band 4) - Philosophische Anthropologie und Handlungslehre* (pp. 385-402). Klostermann.
- Russell, S. (2019). *Human compatible. Artificial intelligence and the problem of control*. Viking.
- Sadin, È. (2020). *La inteligencia artificial o el desafío del siglo*. Caja negra.
- Sanmartí, N. (2020). *Evaluar y aprender: un único proceso*. Octaedro.
- Sennett, R. (2009). *El artesano*. Anagrama.
- Serres, M. (2014). *Hominescence*. Le Pommier.
- Simondon, G. (2008). *El modo de existencia de los objetos técnicos*. Prometeo.
- Snell, B. (2007). *El descubrimiento del espíritu. Estudios sobre la génesis del pensamiento europeo en los griegos*. Acantilado.
- Stiegler, B. (2002). *La técnica y el tiempo (vol. II)*. Hiru.
- Stiegler, B. (2016). *Dans la disruption. Comment ne pas devenir fou?* Babel.
- Tegmark, M. (2018). *Vida 3.0. Qué significa ser humano en la era de la inteligencia artificial*. Taurus.
- Uexküll, J. von (2016). *Andanzas por los mundos circundantes de los animales y los hombres*. Cactus.
- Virilio, P. (1998). *La máquina de visión*. Cátedra.
- Wilson, E. O. (2012). *La conquista social de la tierra*. Debate.

