Persona y Cultura, Universidad Católica San Pablo, número 20-21, pp. 133-147

INTELIGENCIA ARTIFICIAL: ASPECTOS ONTOLÓGICOS Y ÉTICO-SOCIALES[1]

Juan José Sanguineti

Presbítero incardinado en la Prelatura del Opus Dei.
Doctor en Filosofía y Letras por la Universidad de Navarra (España).
Profesor emérito de Filosofía del Conocimiento de la Facultad de Filosofía de la Pontificia Universidad de la Santa Cruz (Italia). Actualmente, es profesor en el Instituto de Filosofía de la Universidad Austral (Argentina).

RESUMEN

La inteligencia artificial realiza automáticamente tareas que el ser humano solo puede llevar a cabo utilizando su inteligencia personal. Por eso, se la califica de artificial ("inteligencia" tiene un sentido analógico). Los resultados son asombrosos y superan la capacidad del hombre y, en cierto modo, corresponden a la función racional humana en su dimensión de cálculo (razón calculadora), separada de la función intelectual (*intellectus*). Sin embargo, la inteligencia artificial no piensa, no tiene conciencia y no puede comprender las cosas esenciales. Los riesgos sociales y éticos no están en la inteligencia artificial como tal, sino en la forma en que los humanos la utilizamos.

1 Este texto está inspirado en una conferencia impartida en la Universidad de Piura (Piura, Perú) en el XVIII Coloquio de Filosofía, el 6 de octubre de 2023.

Palabras clave: Inteligencia artificial, pensamiento humano, cálculo, instrumento, riesgo.

PRIMERA APROXIMACIÓN

La inteligencia artificial (IA) está hoy en primer plano en la atención de todo el mundo[2]. Sus avances en los últimos tiempos, y más aún desde hace casi dos años, son espectaculares. Vuelven a plantearse interrogantes que eran ya conocidos desde hace varias décadas, pero en un nuevo contexto social, debido a la difusión universal de esta invención tecnológica que se presenta hoy como prometedora de beneficios incalculables y revolucionarios para la humanidad, y a la vez como una oscura amenaza para la libertad humana e incluso para la supervivencia del *homo sapiens*.

Seguramente lo mejor para orientarnos en esta problemática es meditar, como sugerirían Heidegger y Ortega, en lo que realmente es la IA y en sus campos de aplicación, que hoy son prácticamente todos, porque no hay casi ningún sector de la vida humana que no quede afectado de un modo u otro por la IA (ciencia, trabajo, educación, economía, política, medicina, etc.).

La inteligencia artificial surgió en los años cincuenta del siglo XX en el contexto de la informática o tecnología de la computación. El nombre de inteligencia artificial fue acuñado en 1956 por el informático John McCarthy en una conferencia internacional sobre el tema, en Dartmouth College (New Hampshire, Estados Unidos).

Se la llamó así porque, mediante la computación (procesamiento de información[3]), realizaba tareas como cálculos, solución de problemas, demostración de teoremas, diagnósticos, juegos como el ajedrez, etc., que tradicionalmente corren a cargo de las personas humanas

2 Se pueden consultar sobre esta temática, los siguientes trabajos: George Luder, *Artificial Intelligence*, Addison Wesley, New York 2009; Stuart Russell y Peter Norvig (eds.), *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River (NJ) 2010; Margret Boden, *Mind as Machine*, Oxford University Press, Oxford 2006; también *Inteligencia Artificial*, Turner Pub., Madrid 2017 y *Artificial Intelligence. A Very Short Introduction*, Oxford University Press, Oxford 2018; Manuel Alfonseca, *Inteligencia artificial* en "Diccionario Interdisciplinar Austral", Claudia Vanney, Ignacio Silva y Juan Francisco Franck (eds.), 2016. http://dia.austral.edu.ar/Inteligencia_artificial; Selmer Bringsjord y Naveen Sundar Govindarajulu, *Artificial Intelligence*, "The Stanford Encyclopedia of Philosophy" (Summer 2020), Edward Zalta (ed.). https://plato.stanford.edu/archives/sum2020/entries/artificial-intelligence.

3 Véase Luciano Floridi, *Information*, Oxford University Press, Oxford 2010 y *The Philosophy of Information*, Oxford University Press, Oxford 2011. Además, David Bawden, Lyn Robinson y Luciano Floridi, *Introduction to Information Science*, Facet Pub., Londres 2022.

cuando usan su inteligencia. Tuvo un desarrollo muy amplio a partir de esa fecha y, si bien hubo un periodo en el que sus posibilidades parecían haberse frenado, en las dos últimas décadas se produjo de golpe un enorme avance inesperado, más aún desde el año 2022, con la aparición del ChatGPT y sus diversas versiones.

Tradicionalmente la IA tenía dos grandes líneas que actualmente se unificaron, pero que al principio surgieron como dos formas independientes de computar. Una es la llamada "arquitectura de Turing" o línea simbólica de computación, basada en programas, símbolos, reglas y algoritmos. La otra línea de computación se basaba en las redes neuronales. Al principio parecían dos modos distintos de computar, uno simbólico y otro más bien asociativo, siendo este último más parecido al modo en que trabaja el cerebro (de ahí el nombre de redes neurales artificiales). En otros tiempos la línea de Turing se consideraba secuencial, mientras que la de las redes neurales trabajaba en paralelo. Esto es así en parte, pero en lo que va del siglo XXI estas dos formas de computar se han fusionado, es decir, la computación simbólica asimiló la elaboración en redes y el paralelismo, y esto es lo que hoy es, sin más, la inteligencia artificial.

Las tareas de las que la inteligencia artificial se ocupa se fueron ampliando cada vez más: deducción, inducción, traducciones, planificación, predicciones, diagnósticos y terapia, lenguaje natural, percepción, resumen de textos, aprendizaje, invenciones artísticas, conducción autónoma de medios de transporte, preparación y difusión de noticias, toma de decisiones. La IA está presente siempre que haya un campo de datos que organizar, de los que se infiere algún resultado (nuevas informaciones, creaciones, planes, trabajos, cálculos), de modo que así puede guiar el trabajo humano, o incluso hacerlo ella misma por su cuenta de un modo increíblemente más veloz y eficiente.

A veces se distingue entre IA *débil*, limitada a la realización de tareas específicas, y *fuerte*, que igualaría e incluso superaría a la inteligencia humana, y que eventualmente sería *general*, como es la humana, es decir, no limitada a cierto tipo de operaciones y ámbitos. Esta distinción es algo relativa y de alguna manera queda superada ante los avances prodigiosos de la IA. En cualquier caso, en este artículo asumo la IA tal como hoy existe. La IA *fuerte* se refiere a especulaciones sobre lo que podría suceder en el futuro, cuando la IA supere la inteligencia de los humanos, como sostiene el transhumanismo. Lo que hoy existe puede considerarse IA *débil*, pero es una IA *débil* que cada vez va siendo más *fuerte*. Sin embargo, como veremos, la IA supera el entendimiento humano en algunas cuestiones, pero no en todas.

INTELIGENCIA RACIONAL Y CÁLCULO

Ante todo, la palabra *inteligencia* es analógica. Hablamos de una inteligencia animal e incluso de una inteligencia natural o en la naturaleza cuando las cosas naturales se comportan "inteligentemente" (con orden, sea sucesivo o con coordinaciones simultáneas, con dosificaciones espacio-temporales apropiadas) y tienen procesos teleológicos, como suelen ser los procesos biológicos, que son útiles para el organismo y las especies, incluso con adaptación al medio y «solución de problemas» si hay obstáculos que evitar, tanto endógenos como exógenos.

Recordemos que la captación de que en la naturaleza y en las leyes naturales físicas hay cierta "inteligencia", como si el orden maravilloso que manifiestan hubiera sido instituido por un agente intelectual, fue tradicionalmente la base para la demostración de la existencia de Dios como Inteligencia creadora (la célebre quinta vía tomista). El comportamiento de las abejas o de las hormigas, con sus niveles organizativos "sociales" tan bien coordinados, hace pensar en una inteligencia (inmanente o trascendente).

En realidad, la IA no comprende nada, no entiende nada, sino que más bien analiza, calcula, deduce. Es decir, podría situarse en lo que tradicionalmente se llama, desde la época de Aristóteles, razón *(logos)*, pero no *intellectus* (comprensión intelectual).

Por eso, cuando el hombre inventa una máquina capaz de realizar tareas "inteligentes", se habla de inteligencia artificial. La computadora no es un organismo, sino un artefacto. Su "inteligencia" no es natural, ni personal, ni biológica, sino artificial y automática. La computadora es una máquina cuyo "trabajo" es la elaboración de la información (procesamiento de datos, secuencias que obtienen resultados buscados). Como toda máquina, la computadora es un instrumento, un artefacto que sirve al hombre para que realice mediante este eso que quizá no quiere o no puede realizar "manualmente", en este caso con su inteligencia natural personal.

Nos sirve porque nosotros sí entendemos la realidad y así incorporamos los resultados de la IA a la realidad entendida. Un mundo con solo IA, sin inteligencias personales, sería vacío, nadie lo entendería (aunque controlara todo el universo), como lo sería un mundo con solo libros, pero sin lectores.

La computación (la palabra originalmente significa "calcular", "contar") consiste básicamente en partir de datos, procesarlos según ciertas reglas y pasos, guardarlos (memoria), sacar de allí resultados y a su vez guardarlos. Este es el esquema habitual del proceso computacional. En rigor sería innecesario decir que este proceso es "inteligente" (la computación lo es de suyo, en cierto modo), pero cuando la IA ejecuta tareas que antes creíamos que solo la inteligencia humana podría realizar (por ejemplo, traducir, resumir, ganar al ajedrez), entonces la hemos llamado "inteligencia artificial". También las viejas calculadoras de los comercios hacían cálculos, pero como esas tareas eran más sencillas, no nos parecía que fueran inteligentes. Aquí siempre estamos jugando con la analogía de la palabra "inteligencia".

En realidad, la IA no comprende nada, no entiende nada, sino que más bien analiza, calcula, deduce. Es decir, podría situarse en lo que tradicionalmente se llama, desde la época de Aristóteles, razón (*logos)*, pero no *intellectus* (comprensión intelectual). El campo en el que la IA trabaja es la racionalidad, con entradas, operaciones y salidas, de modo automático. Se sitúa dentro de una racionalidad que podríamos llamar "de cálculo" (precisamente esto es la computación), y por eso puede trabajar con premisas y hacer operaciones deductivas (sacar consecuencias, por reglas, o por haber "aprendido" por asociación), y así también puede trabajar en función de fines puestos por el programador humano.

Se podría decir, entonces, que la IA es una extensión de la racionalidad humana, o mejor que es una extensión del ámbito del cálculo, el cual no es toda la racionalidad humana, porque realiza de modo automático, sin pensar, eso que nosotros hacemos razonando, pero con pensamiento, percepción y emoción. Además, la IA realiza otras operaciones que simulan la percepción, en el sentido de que interpreta datos sensoriales, identificando en ellos ciertos patrones o con-

figuraciones, por ejemplo reconocer rostros, interpretar un lenguaje (y así podemos "conversar" con ella), o reconocer los signos de una enfermedad, o percibir e identificar música, objetos varios, etc., así como modificar esas objetividades conseguidas o inventarlas a partir de algo dado (por ejemplo, crear música, componer poemas, etc.). Por tanto, es también como una extensión de la potencia perceptiva humana (y animal) e incluso de la imaginación constructiva y por supuesto de la memoria. Todo esto lo hace automáticamente, es decir, separándose de la vivencia psíquica consciente o inconsciente o, dicho de otro modo, del soporte vital —biológico— y más aún del soporte cognitivo animal y humano. No piensa, no siente, no vive, pero consigue resultados que pueden ponerse al servicio de la vida, la conciencia, el pensamiento.

En cierto modo, la IA aprende, al principio, imitando lo que hacen los seres humanos, pero después puede ir a más.

Como todo instrumento, la IA llega más lejos de lo que podemos hacer nosotros intuitivamente o con nuestras propias fuerzas naturales, por lo menos en ciertos ámbitos, así como un automóvil se desplaza mucho más velozmente que nosotros. Personalmente no podemos realizar de modo intuitivo ciertos cálculos, pero la IA sí puede hacerlo, y en este sentido nos sirve, como todo instrumento. Pero nos sirve porque nosotros sí entendemos la realidad y así incorporamos los resultados de la IA a la realidad entendida. Un mundo con solo IA, sin inteligencias personales, sería vacío, nadie lo entendería (aunque controlara todo el universo), como lo sería un mundo con solo libros, pero sin lectores.

Así es como la IA puede vencer a jugadores de ajedrez, como máquina y no como organismo, ni como mente consciente y realmente inteligente. La computadora es máquina porque está hecha en base a un ensamblaje de partes que, unidas, pueden implementar ciertas tareas ("trabajo"). Esto es precisamente una máquina, de donde viene la palabra "mecánico" y "mecanismo".

Lo normal, entonces, es que la IA se vea como un instrumento, como toda máquina, aunque hay que repetir, es una máquina que

amplía el radio de nuestra capacidad operativa (calculística, computacional). Entonces nada tiene de extraño que el modo en que opera se parezca a lo que hacen nuestros procesos mentales o nuestro cerebro. Ya había señalado Aristóteles que *el arte imita a la naturaleza* y, de hecho, nosotros hemos aprendido a fabricar la inteligencia artificial imitando lo que hace el cerebro, en el caso de las redes neurales, o imitando nuestros procesos lógicos mentales, en el caso de la computación simbólica de la línea de Turing. Así también la IA "memoriza" y "aprende" (mejora sus prestaciones en base a lo que hizo antes y no salió tan bien, en función de cierto fin), imitando también aquí lo que hace un organismo cognitivo o solo vegetativo.

En el caso de la computación simbólica (basada en el simbolismo: línea de Turing), se imita lo que nosotros hacemos cuando calculamos deductivamente. En cierto modo, la IA aprende, al principio, imitando lo que hacen los seres humanos, pero después puede ir a más. Por ejemplo: aprende a jugar ajedrez basándose en lo que hacen los agentes humanos. Eso lo va incorporando, lo va mejorando, y así va más allá y puede vencer a un jugador humano de ajedrez y cosas semejantes.

PENSAMIENTO PERSONAL E INTELIGENCIA ARTIFICIAL

Paso a continuación al siguiente punto. ¿Qué tiene el pensamiento personal que no tenga la inteligencia artificial? Estamos aquí ante preguntas filosóficas: ¿En qué se diferencia la IA de la inteligencia personal? ¿Pueden las máquinas pensar? Esta es la problemática introducida ya en 1950 en un conocido artículo de Turing[4]. ¿O no será que el cerebro y nuestro aparente pensamiento no es más que computación, como suelen sostener los materialistas o reduccionistas? Así se pone en tela de juicio —este es el planteamiento del célebre test de Turing— nuestra supremacía como seres inteligentes.

Aquí en realidad hay dos problemas. Uno es ontológico-epistemológico y otro es ético-social, incluso político y educativo. Este último consiste en sentir temor porque la IA puede amenazar nuestra existencia dado que, al superarnos en todos los terrenos, al final se haría más y más autónoma y podría dominarnos. Se verá en el siguiente apartado.

4 Alan Turing, *Computing Machinery and Intelligence*, "Mind", 59 (1950), pp. 433-460. Véase, sobre este punto, mis trabajos *Filosofía de la mente*, Palabra, Madrid 2007, pp. 305-351, y *Ciencia, Tecnología y Mundo humano*, Logos, Rosario 2021, pp. 253-301.

El primer problema se refiere a la distinción esencial entre la inteligencia personal y la IA. Por falta de espacio, no puedo argumentar aquí esta cuestión como sería necesario. Señalo pues intuitivamente, y basándome en la experiencia y el análisis de las operaciones, que la IA realmente no piensa. Opera sin comprensión, no tiene conciencia, no tiene sensibilidad, carece de emociones. Aristotélicamente, podríamos decir que no ejerce operaciones inmanentes. No es un animal, no es ni siquiera un viviente, ni mucho menos una persona. Algunos se aventuraron a decir que un día podría tener conciencia, o que los ingenieros podrían introducirle una conciencia, pero eso es una confusión y es no saber lo que se está diciendo. Sería como creer que cuando la computadora "dice": «necesito una reparación, porque algo mío funciona mal», tendríamos aquí un índice de que tiene conciencia. Pero no es así. Esto es solo una simulación de ser conscientes, salvo que tengamos una idea mecanicista o reduccionista de la conciencia, en la que desaparece la vivencia subjetiva, la experiencia del yo personal, o al menos el yo fenomenológico.

La IA puede simular todo tipo de operaciones vitales, conscientes, animales, humanas. Puede simular emociones, comportamientos personales, un yo. Esta simulación se relaciona con la tesis de los behavioristas radicales de que no habría actos interiores, sino que existirían solamente procesos o fenómenos exteriores. Si es así, si no hay realmente un yo interior, entonces el "yo" ficticio de la IA, reducido a operaciones externas, es absolutamente igual a nuestro pretendido yo interior, que sería igualmente ficticio. La IA podría verse, en este sentido, como una especie de "máquina behaviorista".

La IA realmente no piensa. Opera sin comprensión, no tiene conciencia, no tiene sensibilidad, carece de emociones. Aristotélicamente, podríamos decir que no ejerce operaciones inmanentes. No es un animal, no es ni siquiera un viviente, ni mucho menos una persona.

Pero la IA obtiene resultados reales, como señalé. Aunque simula, *emula*. No razona pensando, pero llega a resultados racionales importantes para nosotros y que nos pueden ser muy útiles, a los que no po-

demos acceder pensando (incluso nos ahorra, en pocos instantes, quizá años de pensamiento humano "calculístico"). Cómo lo hace puede parecer un misterio, pero en lo que expuse anteriormente ya está dada la respuesta. Lo hace con algoritmos, que son procedimientos automáticos, como cuando nosotros realizamos sumas y multiplicaciones siguiendo ciegamente reglas, sin intuiciones eidéticas. También hay algoritmos heurísticos de la IA en los que cabe un aprendizaje, un cierto comportamiento cibernético, pero no en el sentido biológico, sino, como siempre, de modo automático.

¿Tiene esto un límite? ¿Hay algo que la IA no pueda hacer? A veces se han puesto límites falsos a la IA, como que «un gran jugador de ajedrez, con mucha intuición, al final ganaría a la IA». Esto actualmente es poco creíble. En el campo del cálculo, como sacar implicaciones o descubrir *patterns*, no hay límites. El límite fundamental está en que la IA *no comprende*. ¿En qué se ve esto? En que la IA no puede captar finalidades, ni el valor de las personas, o el significado del mundo, o conocer lo esencial de las cosas. Por eso no puede decirnos si Dios existe o no, o cuál es la verdadera religión, o enseñarnos la verdadera moral, o permitirnos dar un valor a la persona o conocer los derechos humanos. Aunque sí podemos programar a una IA, "poniéndole" una ética, una religión, una filosofía, etc. En este sentido, una IA podría personificar a Hegel, por ejemplo, y podríamos charlar con ella como si habláramos con Hegel, lo cual heurísticamente podría tener alguna utilidad.

¿Hay algo que la IA no pueda hacer? El límite fundamental está en que la IA *no comprende*. ¿En qué se ve esto? En que la IA no puede captar finalidades, ni el valor de las personas, o el significado del mundo, o conocer lo esencial de las cosas.

Pero una IA nunca haría grandes descubrimientos o invenciones que requieren una gran creatividad (algo semejante a lo que hicieron Newton o Einstein), y nunca podría decirnos si Hegel está equivocado o no, es decir, no es capaz de hacer una reflexión evaluativa. Tengamos en cuenta, sin embargo, que en nuestro modo de pensar tenemos aspectos automáticos, tanto en las inferencias como en las

asociaciones. La IA nos emula en estos aspectos, pero en ella están desgajados de la comprensión esencial y personal que nosotros, en cambio, asociamos a esos automatismos[5].

Parecería entonces que la IA no podría llegar a pensamientos extraordinarios, revolucionarios, pero que sí ocuparía todo el espacio de nuestra vida ordinaria. En realidad no es así, porque en nuestra vida ordinaria hay muchas cuestiones en las que tenemos que comprender o intuir personalmente, captar valores, tomar decisiones con libertad, cosa que no puede hacer la IA. Captamos personalmente una infinidad de matices contextuales ligados a un fondo de comprensión esencial, mucho más en el trato con personas, un fondo limitado pero real, a lo que no llega la IA sino parcialmente, salvo que sus resultados sean evaluados personalmente y eventualmente corregidos.

Algunos sostuvieron que llegaría un momento en que una IA conseguiría superar el test de Turing (un humano no podría distinguir si está interactuando —conversando— con una IA, por ejemplo un chatbot, o con una persona). En realidad, el hecho de que alguien no pueda distinguir si está charlando con una persona humana o con una IA no es decisivo. Normalmente esto es contextual, porque una persona poco experta podría ser fácilmente engañada durante mucho tiempo y, en cambio, un experto o muy inteligente se daría cuenta al poco tiempo de que está ante una máquina, lo mismo que ahora podemos darnos cuenta de esto con el chatGPT en cuanto lo usamos con mucha frecuencia. A medida que la IA mejora, en algunas competiciones (por ejemplo, en el premio Loebner) las máquinas consiguen superar el test de Turing (engañan a jueces o evaluadores), pero esto en juegos con tiempos limitados.

De todos modos, si no podemos distinguir algo en casos extremos, resulta simplemente que somos engañados, así como hay falsificaciones que quizá no podemos superar. Pensar que ese engaño podría ser universal nos llevaría a un mundo de ciencia-ficción y al problema del "demonio cartesiano", lo que conduce al escepticismo gnoseológico. El problema ontológico del test de Turing, sin embargo, tiene un

5 Señalo dos libros, entre tantos estudios, que ilustran los límites conceptuales de la IA. El primero es el texto de Jobst Landgrebe y Barry Smith, *Why Machines will never rule the World*, Routledge, New York 2023, donde se argumenta que los modelos matemáticos, en los que se basa siempre la IA, no pueden hacerse cargo plenamente ni siquiera de la complejidad biológica, y por tanto no podrían nunca emular la capacidad de la inteligencia humana de gestionar los sistemas dinámicos complejos. Por su parte, Erik Larson en su libro *The Myth of Artificial Intelligence* (Harvard University Press, Cambridge (Mass.) 2021), sostiene que la IA puede hacer inferencias de tipo deductivo e inductivo, pero no abducciones, es decir, adivinar o imaginar hipótesis explicativas (en el sentido de Pierce).

sentido práctico y al final podría ser insuperable si somos behavioristas radicalizados. Pienso que, con un trato prolongado, aparte de exámenes morfológicos, podemos darnos cuenta de si hablamos con *alguien* que tiene una interioridad verdadera (aunque en teoría esto puede ser siempre imitable exteriormente).

RIESGOS SOCIALES Y ÉTICOS

El segundo problema que plantea la IA es ético-social e incluye muchos aspectos[6]. El más dramático, casi apocalíptico, llamado a veces "riesgo existencial", está en pensar que la IA eventualmente podría independizarse y superar tanto a nuestra inteligencia, que el género humano quedaría reducido a una condición marginal en la Tierra y poco a poco se extinguiría[7]. Y aunque podamos "desenchufar" la máquina dotada de IA, en realidad, esta podría aprender a hacerlo por su propia cuenta y buscar las fuentes de energía necesarias para poder así mantenerse activada y reproducirse, mejorando una y otra vez sus versiones. En teoría podría hacer todo eso y, ¿por qué no?, dominar las galaxias.

El riesgo ético actual, más real y más próximo, incluso ya en acto parcialmente, está en que nuestras vidas podrían quedar delegadas, sofocadas o automatizadas por la IA, tanto a nivel individual como social, y eso por nuestra culpa, por falta de virtudes, o por apatía, pero también por una suerte de constricción social que poco a poco se haría inevitable y universal.

Este riesgo por ahora parece remoto, aunque los transhumanistas o futuristas creen en él, incluso como inminente, y algunos se preocupan de obviarlo de un modo u otro (Nick Bostrom, Elon

6 Véase Luc Julia, *L'intelligence artificielle n'existe pas*, Ed. First, París 2011; Luciano Floridi, *The Fourth Revolution*, Oxford University Press, Oxford 2014; Thomas Powers (ed.), *Philosophy and Computing*, Springer, Newark 2017; Mariano Sigman y Santiago Bilinkis, *Artificial. La nueva inteligencia y el contorno de lo humano*, Debate, Buenos Aires 2023.
7 Nick Bostrom, *Superintelligence*, Oxford University Press, Oxford 2014.

Musk, Yuval Noah Harari), así como otros piensan que el hombre, si toma medidas oportunas, podría siempre orientar el desarrollo y las aplicaciones de la IA, la cual de todos modos tiene los límites conceptuales indicados arriba. Es decir, si fuéramos dominados por las IA tal como vaticinan algunos, no se produciría un real mejoramiento de la inteligencia, sino solo un super-dominio técnico universal pero vacío, puramente calculístico. Se trataría de post-humanos que no serían más que máquinas impersonales[8].

En un plano más directamente personal, el principal peligro está en llegar a una renuncia a pensar en las cuestiones esenciales, religiosas, metafísicas, personales, y a delegar nuestras decisiones más importantes en la IA. Esto equivaldría a abandonar el *intellectus* y caer en la pura razón calculadora, basada por ejemplo solo en estadísticas. Quedaríamos así cosificados, reducidos a objetos. Si la IA lo resuelve todo, no nos queda casi nada por pensar y decidir. Esta amenaza puede evitarse si aprendemos a utilizar la IA como un instrumento técnico importante pero siempre auxiliar, contando para ello con virtudes intelectuales y morales.

El riesgo ético actual, más real y más próximo, incluso ya en acto parcialmente, está en que nuestras vidas podrían quedar delegadas, sofocadas o automatizadas por la IA, tanto a nivel individual como social, y eso por nuestra culpa, por falta de virtudes, o por apatía, pero también por una suerte de constricción social que poco a poco se haría inevitable y universal, tanto en sociedades totalitarias (allí mucho más, evidentemente), como en las sociedades supuestamente liberales. Los márgenes de libertad de nuestras vidas se irían encogiendo cada vez más, o quedarían reducidos a cosas triviales. Como

8 Dejo de lado otros aspectos considerados por el transhumanismo, ya que no es este el objetivo de este artículo.

hace notar Shannon Vallor, «lo que la IA pone en mayor peligro no es la mera supervivencia de nuestra especie, sino nuestra propia autocomprensión de la perpetua libertad y responsabilidad»[9].

Este riesgo lo tiene siempre el uso de la tecnología automatizada, sobre todo cuando asume un desarrollo abrumador, que aquí amenaza con ocupar todos nuestros espacios de creatividad, libertad y pensamiento. La total tecnificación de la vida despersonaliza. La IA hace más amenazante este peligro a causa de su actual crecimiento exponencial y su penetración y aplicación en todos los ámbitos.

En un plano más directamente personal, el principal peligro está en llegar a una renuncia a pensar en las cuestiones esenciales, religiosas, metafísicas, personales, y a delegar nuestras decisiones más importantes en la IA. Esto equivaldría a abandonar el *intellectus* y caer en la pura razón calculadora, basada por ejemplo solo en estadísticas. Quedaríamos así cosificados, reducidos a objetos. Si la IA lo resuelve todo, no nos queda casi nada por pensar y decidir.

Creo que esta amenaza puede evitarse si aprendemos a utilizar la IA como un instrumento técnico importante pero siempre auxiliar, contando para ello con virtudes intelectuales y morales[10], como la curiosidad, la visión crítica, saber hacerse preguntas, pensar por cuenta propia, moderación, tratar de comprender lo esencial, conocer mejor a las personas, veracidad, flexibilidad, apertura de pensamiento, fomentar las relaciones personales, prudencia, sin "digitalizar" toda nuestra vida[11].

La IA es como una "objetividad abstracta" e impersonal, aunque simule interacciones personales, que debe intervenir en nuestra vida personal en la medida en que es realmente útil. Ciertamente la IA tiene más utilidad en cuestiones técnicas, organizativas, logísticas, matemáticas, y menos en cuestiones humanas, donde puede ser más bien un auxiliar, un consejero, pero nada más. La IA tiene menos protagonismo en la educación, en la familia, en las amistades, en la religión.

En el mundo del trabajo, en las empresas, en la economía, en la medicina, en la política, la IA sirve siempre que la utilicemos sin sustituir las relaciones personales y contando con los criterios prudenciales que solo la inteligencia personal posee, aun cuando esta sea falible y limitada.

9 Shannon Vallor, *The AI Mirror*, Oxford University Press, Oxford 2024, p. 195.
10 Véase Shannon Vallor, *Technology and the Virtues*, Oxford University Press, Oxford 2016.
11 Véase mi trabajo *Virtudes intelectuales*, Diccionario Interdisciplinar Austral, Claudia Vanney, Ignacio Silva y Juan Francisco Franck (eds.), http://dia.austral.edu.ar/Virtudes_intelectuales, octubre de 2023.

Las personas llegan a captar realidades inaccesibles a la IA porque captan aspectos humanos, éticos, emocionales, relacionales, personales, a los que no llega una no-persona. No tiene sentido ser amigo de una no-persona. En la "amistad" con un robot humanoide con IA nos estamos haciendo amigos de una realidad impersonal.

Todo lo que obstaculice las relaciones personales no es deseable. Sin duda que esto a veces se impone inevitablemente a causa de la masificación, que nos obliga a tratar con colectividades en las que los criterios se vuelven estadísticos, matemáticos, como si las personas y sus problemas fueran números. Pero se trata de organizarnos de modo que esto no suceda, para lo cual hace falta ver a las personas en su realidad concreta y viva, profunda, irrepetible, inabarcable con el conocimiento que puede dar de ella una IA.

Si queremos superar los riesgos sociales de la IA, tenemos que mejorar nosotros, porque la actual IA es un reflejo de lo que hoy somos nosotros.

Los riesgos sociales concretos de una IA no debidamente controlada y orientada al bien, al servicio del hombre y de la sociedad, hoy son ampliamente conocidos (sesgos, no privacidad, pérdida de control y tantos otros). La IA implica un poder inmenso, autónomo, automático y creciente. Todo poder puede ser utilizado para el mal, para el delito, para la guerra, para engañar. Esto siempre fue así. Puede ser usado para el mal, o causar daños "irresponsables" precisamente por su automatismo y por las consecuencias colaterales imprevisibles y no deseables derivadas de ese automatismo.

La IA la hacemos los seres humanos. Tal como está ahora, con la selección de datos que opera y los algoritmos que de ahí resultan, ella refleja lo que de modo consciente o no diseñan sus creadores, con sus sesgos, posibles discriminaciones y limitaciones, por ejemplo, si detrás hay prioridades comerciales o económicas, o intereses educativos, científicos o de otro tipo. Si queremos superar los riesgos sociales de la IA, tenemos que mejorar nosotros, porque la actual IA es un reflejo de lo que hoy somos nosotros[12]. La técnica siempre tiene que estar guiada por la ética. Pero el riesgo está también en contar

12 Véase Shannon Vallor, *The AI Mirror*, ob. cit.

con una "ética" equivocada (individualista, hedonista, etc.), lo cual reside no en la IA, sino en las personas humanas, pero que puede incorporarse a la IA.

El verdadero riesgo no viene de la IA, que puede ser un instrumento maravilloso para el pensamiento y el trabajo, sino del uso que nosotros hagamos de ella, que a veces depende de nosotros personalmente, y a veces de que nos lo impongan en cuanto usuarios las estructuras sociales, políticas, empresariales, económicas. Por ejemplo, si nuestra salud fuera a quedar en un futuro en manos de la IA y no de médicos ni enfermeros humanos, el riesgo es serio, pero además tenemos que saber por qué y en qué, según los argumentos presentados en este artículo. Es obvio que la IA puede superar enormemente la competencia de médicos, en diagnósticos y terapia, pero no en todos los aspectos (personalización, contextualidad).

Estamos en una etapa de la civilización en la que la digitalización supone un avance tecnológico revolucionario en la historia de la humanidad. La IA nos está dando un dominio inédito sobre el universo (aunque no total). Podemos servimos de él con prudencia y sabiduría personal, una prudencia también política, con la exigencia de regulaciones sociales (las cuales se pueden hacer bien o mal) ancladas en virtudes y valores. El núcleo del dominio sapiencial y prudente de la IA es el bien de la persona.

Hoy somos más conscientes de los riesgos que supone contar con un instrumento tan valioso y poderoso como la IA. Nuestra actitud ante la tecnologización de la inteligencia del cálculo tiene que ser positiva, siempre que recordemos los límites de tal inteligencia y siempre que sepamos mirar la realidad y las personas con la visión que da la inteligencia personal.