De la inteligencia artificial como instrumento: la tesis de la dependencia cognitiva

On Artificial Intelligence as an Instrument: The Cognitive Dependency Argument

Marcos de J. Aguirre Franco
10.26754/ojs_arif/arif.202419489

Resumen

Por diversas razones, los filósofos Hubert Dreyfus y John Searle no están de acuerdo en la posibilidad de que la Inteligencia Artificial (IA) sea capaz de pensar y comprender como la hacen los seres humanos. Tanto el enactivismo de Dreyfus (1992) como la posición semántica de Searle (1980) sirven de base para la tesis de la dependencia cognitiva que aquí se defiende, la cual considera que la IA es incapaz de desarrollar sus funciones sin el sostén de las capacidades cognitivas que se derivan de la experiencia y creatividad de los seres vivos, en especial de la vida humana.

Palabras clave: Inteligencia artificial, enactivismo, semántica, representación simbólica, dependencia cognitiva

Abstract

For various reasons, the philosophers Hubert Dreyfus and John Searle disagree on the possibility that AI is capable of thinking and understanding as humans do. Both the enactivism of Dreyfus (1992) and the semantic position of Searle (1980) serve as the basis for the thesis of cognitive dependence defended here, which considers that AI is incapable of developing its own functions without the support of the cognitive abilities that derive from the creativity and experience of living beings, especially human life.

Keywords: Artificial intelligence, enactivism, semantics, symbolic representation, cognitive dependence

Recibido: 3/7/2023. Aceptado: 15/11/2023
Análisis. Revista de investigación filosófica, vol. 11, n.º 1 (2024): 63-79 ISSNe: 2386-8066

1. Introducción al argumento de la enacción en Dreyfus

En los últimos años, el desarrollo de la inteligencia artificial (IA) ha tenido un auge y desarrollo significativo, una razón que en cierto modo ha conducido a científicos y filósofos a reafirmar su posición con respecto a la idea de si la IA ha podido ya superar la prueba de Turing o incluso a considerar si su actual capacidad de procesamiento ha podido conducir hacia una forma genuina de pensamiento.

Ya en la década de los noventas del siglo pasado, el filósofo Hubert Dreyfus (1992) insistía en que el sentido común, tal como lo conocemos, no puede ser considerado como un atributo procedimental de la IA cuando se tiene en cuenta que el sentido común depende de un proceso *enactivo*[1] propio de aquellos sistemas biológicos que se desarrollan en función de un entorno que les es *coextensivo*[2], esto es, del espacio vital que otorga la contención necesaria para la toma de decisiones que influyen en la constitución biológica de los organismos. Más recientemente y en continuidad a la postura de Dreyfus (1992), el profesor Tim Crane sugirió que una de las razones por las cuales «el conocimiento de sentido común no puede ser representado como un manojo de reglas y representaciones es que el conocimiento de sentido común es, o cuenta con serlo, una especie de *saber-cómo*» (Crane, 2022: 197). Según él, algunos investigadores cognitivistas se han visto obligados a distinguir el saber que algo es el caso y el saber cómo hacer algo. El *saber-que* por ejemplo, se refiere a la adquisición de cierto tipo de conocimiento enfocado hacia cierta clase de «hechos» que de alguna manera han sido aprobados o institucionalizados de manera colectiva, como por ejemplo, el «hecho» de que tal o cual edificio funciona para tal o cual uso; que un determinado animal entra

[1] Enactivismo. «Corriente en ciencias cognitivas contraria al computacionalismo. El enactivismo interpreta toda actividad cognitiva como radicada en la interacción dinámica entre un organismo y su entorno» (Arias, 2021: 338).

[2] La idea de coextensividad aquí propuesta, implica considerar que el entorno se extiende a los organismos (biotopo-biocenosis) de la misma manera que los organismos se extienden a su entorno. En este sentido, el entorno (biotopo) contiene la información que los procesos cognitivos de los organismos (biocenosis) pueden llegar a captar e interpretar (influyendo en aquellas decisiones que en último término conducen a su adaptación, configuración y auto-mantenimiento); y de la misma manera, la cognición que los organismos han alcanzado en la práctica y enacción en el propio entorno, contiene ya la información suficiente para continuar con el proceso de adaptación y auto-mantenimiento. Esta idea se complementa con la noción de *Umwelt* o mundo circundante que fuera desarrollada por el biólogo alemán Jakob von Uexküll (2010) (ver la nota al pie número 11).

en tal o cual categoría taxonómica o que un título universitario acredita a ciertos individuos para que puedan practicar tal o cual profesión.

Por lo que respecta al *saber-cómo*, Crane (2022) considera que ciertas capacidades como por ejemplo montar en bicicleta, no representan un hecho que pueda ser reducido al conocimiento de hechos basados en reglas y principios. Por lo tanto, el *saber-cómo* no requiere el establecimiento rígido de instrucciones que deban llevarse a cabo para que se produzca el «hecho» de montar en bicicleta; la esencia del *saber-cómo* presupone entonces la imposibilidad de acceder a un conocimiento enteramente basado en la razón. Esto es así, ya que montar en bicicleta, por ejemplo, supone un *proceso* experiencial fundamentalmente enactivo.

Es por ello que el aprendizaje de montar en bicicleta no se puede establecer según principios predefinidos del tipo «"al dar vuelta a la derecha en una esquina inclínese ligeramente a la derecha con la bicicleta". Sencillamente se *agarra el modo* mediante un método de prueba y error» (Crane, 2022: 197). Según Crane (2022), Dreyfus (1992) asume que *agarrar el modo* implica una acción que requiere tener acceso a una *inteligencia general*, es decir, al tipo de inteligencia a la que se puede acceder a través de la relación inextricable entre la mente, el cuerpo y el entorno[3].

Así pues y como bien ha sugerido Dreyfus (1992), el conocimiento del *saber-cómo* implica una relación fenomenológica entre un cuerpo biológico y el entorno que le es ineludiblemente coextensivo. Por lo tanto, parece evidente que la IA es incapaz de tener acceso a un tipo de conocimiento *experiencial* ya que para ello tendría que estar habilitada con un cuerpo biológico que permita dar *sentido* y significado a la realidad que es decodificada a partir de un *saber-qué,*[4] es decir, a través de todo el bagaje de conocimiento de hechos que fueron originados en la intrinseca relación entre las impresiones y las ideas.

Por tanto parece difícil proyectar que la IA pueda tener la capacidad de procesar el mundo de manera cognitiva y fenomenológica (es decir, a partir de aquello que por experiencia podemos llamar pensamiento). Para Crane, la posibilidad de acceder a esta capacidad tiene por lo menos dos objeciones filosóficas sobresalientes:

---

[3] Sobre la idea de relación mente-cuerpo-entorno, el profesor Alva Noë ha dedicado una pormenorizada argumentación en su libro *Out of our heads* (2009).

[4] «Saber *qué es una silla* no es nada más una cuestión de saber la definición de la palabra «silla». También implica esencialmente saber qué hacer con las sillas, cómo sentarse en ellas, levantarse de ellas, ser capaz de decir qué objetos de la habitación son sillas, o qué clases de cosas pueden usarse como sillas si no las hay a la mano» (Dreyfus en Crane, 2022: 197).

> 1. Las computadoras no pueden pensar porque pensar requiere capacidades que las computadoras, por su naturaleza misma, no pueden tener nunca. Las computadoras tienen que obedecer reglas (ya sean algorítmicas o heurísticas), pero pensar no puede ser capturado en un sistema de reglas, sin importar cuan complejas sean. Lo que requiere el pensamiento es un compromiso activo con la vida, la participación en una cultura y un «saber cómo» que nunca puede ser formalizado a través de reglas. Este es el enfoque asumido por Hubert Dreyfus en su crítica corrosiva de la IA: *What Computers Can't Do*.[5]
> 2. Las computadoras no pueden pensar porque solo manipulan símbolos de acuerdo con sus rasgos formales; no son sensibles a los significados de estos símbolos. Este es el tema de un argumento bien conocido de John Searle: el «cuarto chino». (Crane, 2022: 192 y 193) (se volverá a él más delante)

Una de las dudas que constantemente genera la IA es la de cómo se podrían adaptar sus decisiones a cierto tipo de eventualidades que tienen especial importancia para sistemas biológicos y culturales dotados con una inteligencia general. Parece obvio decir que las posibles decisiones que un ser humano puede llegar a realizar superan, en complejidad, a las decisiones que, por otro lado, una IA debe seguir si ha de adecuarse a sus propios procedimientos.

Es común encontrar historias que ponen en duda la capacidad de la IA para tomar decisiones correctas ante cierto tipo de eventualidades[6]. Como bien supuso Dreyfus (1992), en tales casos el «sentido común» es clave para actuar con cierta coherencia; sin embargo, el procedimiento que permite hacer frente a las particularidades exige la flexibilidad adaptativa que de hecho es característica de una inteligencia general que se ha desarrollado de una manera biológica y cultural.

---

[5] En 1972 el filósofo Hubert Dreyfus escribió *What Computers Can't Do*, libro que tuvo algunas revisiones posteriores en 1979 y 1992, esta última con un título actualizado, *What Computers Still Can't Do: A critique of artificial reason,* libro publicado por The MIT Press.

[6] Una de esas historias es la que fue publicada en el Arizona Daily Star el 31 de mayo de 1986. En ella se describe un incidente que ha llegado a poner en duda la efectividad de las reglas y procedimientos que han de llevarse a cabo ante una situación particular. El profesor Tim Crane citó esta historia junto a su fuente: «Un conductor de autobus novato, suspendido por no hacer lo indicado cuando una muchacha sufrió un ataque cardiaco en su autobus, seguía reglas perfectamente estrictas que prohíben a los conductores dejar su ruta sin permiso, dijo ayer un funcionario de la unión. «Si hay que acusar a alguien, póngase la acusación en las reglas que estas personas tienen que seguir» (dijo el funcionario). (Un vocero de la compañía de autobuses defendió las reglas:) "Da usted una mínima libertad y ¿dónde acaba la cosa?"» (Crane, 2022: 194).

Sobre esto último queda la duda de si aún está abierta la posibilidad de que algún día la IA (ya sea a través del simbolismo o las redes neuronales) tenga la capacidad de gestionar la complejidad de una realidad que, por naturaleza y cultura, le es significativa al ser humano.

Aunque todavía existen algunas objeciones por parte de quienes defienden que las funciones de la IA involucran pensamiento en un sentido general, estas aún no han sido suficientemente fuertes como para considerar que tales funciones han alcanzado una capacidad de decisión basada en un *saber-cómo,* es decir, en la forma de inteligencia general a la que puede acceder una persona sana en su entorno natural y social.

Además, si se tiene en cuenta que el ser humano y su medio es un *proceso* que se halla en continua constitución evolutiva, se podrá aceptar que la resolución de ciertos dilemas morales, científicos y culturales que le son significativos, requerirán de una capacidad creativa que sea capaz, constantemente, de trascender el aprendizaje de las generalidades que subyacen en los casos particulares surgidos en el pasado.

Teniendo en cuenta las anteriores consideraciones de Dreyfus (1992), será preciso continuar con nuestra argumentación partiendo de la siguiente pregunta: ¿Cómo podría la IA adaptar la generalidad de sus decisiones a circunstancias concretas propias de los estados experienciales, emocionales y racionales de sistemas biológicos como el de los seres humanos que han evolucionado en su entorno? Si una decisión efectiva, además de la inteligencia inductiva y deductiva, precisa de una capacidad para «imaginar» escenarios contrafácticos (que van más allá de la mera generalización del *aprendizaje profundo*[7] basado en semejanzas) el desarrollo de la IA de algún modo debería trascender algunas de sus capacidades

[7] Aprendizaje profundo (*deep learning*) "es un tipo de aprendizaje automático (*machine learning*) e inteligencia artificial que imita la forma en que los humanos obtienen ciertos tipos de conocimiento. […] Los programas informáticos que utilizan el aprendizaje profundo pasan por el mismo proceso que el niño pequeño que aprende a identificar a un perro. Cada algoritmo de la jerarquía aplica una transformación no lineal a su entrada y usa lo que aprende para crear un modelo estadístico como salida. Las iteraciones continúan hasta que la salida alcanza un nivel aceptable de precisión. La cantidad de capas de procesamiento a través de las cuales deben pasar los datos es lo que inspiró la etiqueta *profundo*." Extraído del sitio web *ComputerWeekly.es*. Este texto fue realizado por Ed Burns en septiembre de 2021. Para más información veáse la siguiente liga: https://www.computerweekly.com/es/definicion/Aprendizaje-profundo-deep-learning

procedimentales si lo que se busca con su continuo perfeccionamiento es alcanzar una inteligencia que incluya una forma de *creatividad artificial* (creatividad computacional o creatividad mecánica). Claramente, al día de hoy esto no ha podido justificarse de manera plena en la investigación de la IA ya que muchos de los procesos mentales que le son atribuidos a la creatividad humana son, por su naturaleza única y no-convencional, todavía desconocidos por los propios seres humanos. Ciertamente, esta limitación supone un obstáculo para reproducir, de manera artificial o computacional, soluciones novedosas que en principio han sido derivadas de necesidades estrictamente biológicas como respuesta adaptativa a un entorno de cambios.

Ahora bien, si en algún futuro fuese posible desarrollar una *creatividad artificial* equivalente a la creatividad de aquellos organismos biológicos que enactúan en un entorno que les es evolutivamente coextensivo, ello permitiría que la IA pueda alcanzar el ajuste fino que exigen las situaciones particulares que se originan en la complejidad natural y social de organismos biológicos con necesidades correspondientemente biológicas, es decir, en los nuevos escenarios que exigen respuestas creativas (y adaptativas) a problemas específicos que por razones obvias resultan difíciles de generalizar. Aunque hoy en día esto parece improbable, si en algún futuro fuese posible desarrollar una forma de *creatividad artificial* equivalente a la del ser humano, tal creatividad tendría que estar mediada por la información que continuamente genera (y actualiza) la relación de un cuerpo biológico (con sensaciones biológicas, percepciones biológicas y pensamientos biológicos) en un entorno que le es coextensivo desde el punto de vista evolutivo.[8]

Así pues, para tomar una decisión creativa (*thinking outside the box*) que permita ir más allá de la pura semejanza, el proceso tendría que ser inverso pues antes de que el aprendizaje profundo permita que la IA pueda encontrar patrones (previamente aprendidos) de similitud con los casos particulares, será necesario que la IA sea capaz de aprender de las «diferencias» del entorno que interesan a la biología de cada organismo y que en última instancia alimentan las excepciones a la regla más

[8] Ciertamente, el tipo de creatividad que aquí se aborda va más allá de la creatividad que hoy en día se puede encontrar dentro del paradigma conexionista, un tipo enfoque en el que incluso se han llegado a proponer sistemas dinámicos centrados en emular el funcionamiento biológico de los sistemas neuronales. Aunque aquí se analiza principalmente el paradigma simbólico, la argumentación aquí desarrollada en favor de la «tesis de la dependencia cognitiva» se dirige a cualquier paradigma que defienda la posibilidad de independencia y autonomía de la IA.

que a las semejanzas[9]. Como escribió Manuel Gausa (2000), no se puede seguir sacando brillo a la manzana de siempre. Este punto conduce, por mor al argumento del sentido común, a que se tenga en cuenta el equilibrio entre el hábito de las semejanzas y la creación de las diferencias, una proporción que en todo caso parece haber sido decisiva para la evolución biológica y cultural del ser humano.

Evidentemente, hay casos particulares que requieren de la inteligencia (propia de las semejanzas) y no necesariamente de actos creativos (fundados principalmente en el procesamiento de las excepciones y las diferencias). Más allá de esto, si los principios de la IA justifican la posibilidad de alcanzar una forma de *inteligencia general*, esta deberá tener en cuenta a la creatividad (que ha permitido la continua adaptación biológica al entorno) como parte de sus fundamentos.

Ahora bien, si se sigue la explicación de Crane (2022), cabría entonces preguntar si en los principios del pensamiento que le son atribuidos a la IA se considera la capacidad de romper ciertas reglas y procedimientos que han nacido tanto del aprendizaje automático (*machine learning*) como del profundo (*deep learning*). Si esto es así, para romper ciertas reglas de la programación, primero será necesario que la IA tenga la capacidad de generar necesidades (biológicas) que en principio le son propias a los organismos que se hallan en coextensión a su entorno. Evidentemente, a falta de la complejidad fenomenológica como la de los organismos vivos en su entorno de cambios, la IA será incapaz de generar necesidades que conduzcan a romper las reglas que fueron consideradas en su programación.

En resumen, el *saber-cómo* que le es propio a la *inteligencia general* (biológica) es un requisito ineludible para que la IA pueda alcanzar una forma de conocimiento de sentido común como el que practican los organismos que han evolucionado en su entorno. Sin embargo, el sentido común en una IA es lo que de hecho Dreyfus (1992) cree que es imposible alcanzar. Como él mismo escribió, «la inteligencia

[9] Un aprendizaje a partir de diferencias más que de semejanzas, implica recurrir a la imaginación (mediante hipótesis y (o) contrafácticos), un proceso que no está presente en aquellos casos que han sido aprendidos con anterioridad a partir de similitudes o elementos de generalización. El aprendizaje a partir de «diferencias» implica por tanto diferencias en la manera en que los organismos se aproximan a los nuevos acontecimientos de *su* entorno. La capacidad humana de «virtualización» es un ejemplo de ello. Para el profesor Pierre Lévy «la virtualización puede definirse como el movimiento inverso a la actualización. […] La actualización iba de un problema a una solución. La virtualización pasa de una solución dada a un (otro) problema» (Lévy, 1999: 19-20). Esta capacidad implica la creatividad de un organismo que constantemente debe tomar nuevas decisiones en función de los cambios (y diferencias) que se producen en *su* entorno.

humana requiere el trasfondo de sentido común que los seres humanos adultos poseen en virtud de tener cuerpos, interactuar hábilmente con el mundo material y haber sido adiestrados en una cultura» (Dreyfus en Crane, 2022: 196).

2. Searle y la habitación china

Por otra parte, la argumentación de John Searle (1980) contra la posibilidad de que la IA pueda siquiera aproximarse a una forma de *inteligencia general*, se basa en ciertos principios cognitivos justificados a partir de la semántica. En este sentido, la posición del filósofo estadounidense se aproxima al argumento que se presenta en este artículo. Antes de ir a él, es preciso considerar algunas de las implicaciones más generales que subyacen a la tesis de Searle (1980).

Para esquematizar brevemente su posición filosófica (en contra de la posibilidad de que la IA tenga una capacidad real de pensamiento, o, simplemente para rebatir la prueba de Turing), es muy conocido el experimento mental de la «habitación china» que Searle publicó en 1980 en un artículo que tituló *Mind, brains and programs*.

Para explicar su estructura, en el experimento Searle se imagina dentro de una habitación que tiene dos aberturas que dan al exterior, por ejemplo A y B (entrada y salida de información respectivamente).

Por la abertura A ingresan hojas de papel con complejos sinogramas (caracteres chinos) a manera de preguntas que deben ser respondidas por Searle desde el interior. Sin embargo y para facilitarle la tarea (dado que Searle no sabe chino), dentro de la habitación hay un gran estante con libros en inglés en donde se especifican las instrucciones (reglas y procedimientos) de respuesta que deben seguirse cuando ingresan determinados tipos de marcas inscritas (sinogramas) en las hojas de papel. Estas instrucciones indican la hoja a elegir (de entre un montón de hojas que se hallan dispersas en la habitación), la cual tiene inscritos aquellos sinogramas correctos que deben ser enviados como respuesta hacia el exterior de la habitación por la abertura B. Este intercambio permite que la persona del exterior que habla el idioma chino pueda comunicarse eficientemente con el interior de la habitación, ello, a pesar de que allí dentro no se conozca el idioma chino.

De manera muy simplificada, el experimento mental intenta ejemplificar, como supone el propio Searle (1980), lo que ocurre cuando se ingresa cierto tipo de información en un ordenador o computadora.

Para Crane, el experimento de la habitación china «muestra que hacer funcionar un programa de computadora nunca puede constituir un genuino entendimiento o

pensamiento, ya que todo lo que las computadoras pueden hacer es manipular símbolos de acuerdo con su forma» (Crane, 2022: 203). Que la capacidad de respuesta de una computadora precise de cierta eficiencia, ello, según Searle (1980), no es una propiedad suficiente para que se produzca una comprensión real a la manera de una Inteligencia Artificial Fuerte (IAF).

Aunque existen conocidas objeciones a la argumentación de Searle (1980) su tesis ha podido resumirse en cuatro puntos:

> 1. Los programas de computadora son puramente formales o «sintácticos»: a grandes rasgos son sensibles únicamente a las «figuras» de los símbolos que procesan.
> 2. La compresión genuina (y, por extensión, todo pensamiento) es sensible al significado (o «semántica») de los símbolos.
> 3. La forma (o sintaxis) nunca puede constituir, ni ser suficiente para ello, un significado (o semántica).
> 4. Por lo tanto, hacer funcionar un programa de computadoras nunca puede ser suficiente para la comprensión o el pensamiento. (Crane, 2022: 203)

Una conocida crítica a esta argumentación es justamente la simplicidad con la que Searle (1980) aborda la cuestión. Según sus detractores, Searle (1980) hace trampa en su argumento cuando admite que las computadoras son incapaces de «comprender» el idioma chino solo porque en su analogía *él* es incapaz de comprender el chino.

> Pero sus críticos responden que esto no es lo que debiera decir la IA. Searle-en-el-cuarto es análogo a sólo una parte de la computadora, no la computadora misma. La computadora misma es análoga a Searle + el cuarto + los trozos de papel (los datos). Así, afirman los críticos, Searle está proponiendo que la IA pretende que una computadora comprende porque una parte de ella comprende; pero nadie trabajando sobre la IA diría eso. Antes bien, dirían que todo el cuarto (o sea toda la computadora) entiende chino. (Crane, 2022: 204)

Según Crane (2022), la objeción de Searle a este argumento es que aún si se pudiera memorizar la totalidad de las instrucciones (es decir, las reglas, los procedimientos y los datos de cada símbolo o sinograma) para hacer lo mismo que se hizo en la habitación pero ahora fuera de ella, seguiría sin haber una comprensión o entendimiento real. En todo caso, según él, habría una «simulación» de un entendimiento característico de una Inteligencia Artificial débil (IAD). A pesar de esto, los detractores aún argumentan que el proceso de memorización de las reglas, procedimientos así como de los datos de cada sinograma, es un proceso normal que una vez realizado inevitablemente conducirá a la comprensión y al

entendimiento tal como hacemos los humanos, esto supone, según afirman, de lo que se trata el aprendizaje de un idioma.

Lo anterior implica que, al salir de la habitación con todo el bagaje producido a través de la memorización de los procedimientos, caracteres, símbolos, etc., ya se estaría generando cierta interacción con el mundo (entorno) lo que podría conducir a una forma de interacción en el que un determinado símbolo o sinograma tendría su correlato en el mundo que ahora es exterior a la habitación china. Pero de esto trata justamente la argumentación de Searle (1980) con respecto al hecho de que una computadora solo puede «simular» el pensamiento a través de interacciones con el mundo, pero no llegar a realizarlo. En este sentido, si existe una relación entre las representaciones simbólicas y el mundo que permita realizar acciones más o menos «eficientes», para Searle (1980), no es lo mismo que haber alcanzado la comprensión genuina que requiere la experiencia fenoménica. Como se dijo anteriormente, para alcanzar el pensamiento real (o mental), se debe estar en una relación coextensiva entre el cerebro, el cuerpo y el entorno.

Al parecer, lo que intenta expresar la objeción de Searle es lo siguiente: «si dice usted que, a fin de pensar, necesita interactuar con el mundo, entonces habrá abandonado la idea de que una computadora puede pensar sencillamente porque es una computadora» (Crane, 2022: 205). Aunque la IA tenga una gran capacidad para manipular eficientemente los símbolos a razón de ciertas reglas y procedimientos fundamentalmente sintácticos, la interacción con el significado mundo se mantiene como el fundamento de aquello que, por experiencia, podemos realmente llamar pensamiento.

En una palabra, aceptar que la sintaxis es un factor suficiente para que la IA pueda llegar a generar una comprensión real, implica desestimar que el significado sea necesario para «comprender» (lo que significa) un tipo de organización como el de la sintaxis. En último término, la razón de ser de la sintaxis deberá apuntar a las funciones cognitivas que estén capacitadas para las operaciones semánticas como las de los seres humanos que requieren dar sentido a su entorno.

3. La IA como extensión instrumental de la inteligencia general: el argumento de la dependencia cognitiva

Si el reconocimiento de cierto tipo de caracteres o símbolos como posibles respuestas que pueden darse a determinadas preguntas se considera una demostración de que se han «comprendido» tanto las preguntas como las respuestas en el idioma chino, esto, en principio, debería conducir a un error, ya que la

«comprensión» del idioma chino implica haber tenido la «experiencia» —recuérdese la enacción en Dreyfus (1992)— que permite que los referentes simbólicos que han ingresado por la habitación tengan algún sentido o razón de ser en el mundo que se encuentra fuera de ella, esto es, que los referentes simbólicos sean efectivamente representaciones semánticas atribuibles a la *experiencia* del entorno.

Como llegó a sugerir Searle (1980), comprender las formas simbólicas de manera sintáctica no es lo mismo que comprender la realidad de los hechos a los que dichas formas simbólicas se refieren. Es evidente que el lenguaje simbólico no es la realidad de los hechos que ocurren en la experiencia (a pesar de que el lenguaje intenta mantener cierta isomorfía[10]). Esta diferencia nos lleva a suponer que el reconocimiento simbólico (sintaxis) no es el reconocimiento del significado (semántica) que se presenta en la comprensión de la experiencia, ello, a pesar de que la organización sintáctica remita a ella. Ahora bien, dado que el *sentido* o la razón de ser de la IA «depende» de la significación (semántica) que el ser humano le otorga a cierto tipo de símbolos que continuamente manipula y decodifica, la IA, por sí sola, parece incapaz de pensar o comprender el «sentido» de un mundo que se encuentra invariablemente mediado por estados sensoriales, emocionales y valorativos que cambian, no solo en los diferentes contextos espacio-temporales, sino en cada individuo que participa fenomenológicamente en la interpretación de su propio mundo.

Por otro lado, que la IA pueda tener acceso al bagaje de los símbolos geométricos que el ser humano utiliza para reconocer los conceptos (universales) de lo que significa una mesa particular, un árbol, un edificio etc., esto no debería considerarse una especie de indicador de que la IA ha sido capaz de «comprender» el sentido y significado de dichos símbolos geométricos puesto que una mesa, un árbol o un edificio *del mundo* tienen sentido y significado únicamente dentro del entorno adyacente que le es correlativo a la experiencia humana que ha sido enriquecida con un complejo sistema fenomenológico y cognitivo; una experiencia que, por otra parte, tiene la particularidad de actualizar continuamente el significado de los objetos del mundo en función de las relaciones factuales que es capaz de producir un sujeto anímico dentro de un grupo social en determinadas circunstancias. Este proceso

[10] El término isomorfía es usado aquí según la teoría del lenguaje del primer Wittgenstein: «La tesis del isomorfismo establece que para que el lenguaje pueda representar la realidad tiene que haber un mínimo común idéntico, y ese mínimo común es precisamente la forma lógica, el modo y manera en que los elementos del lenguaje, por un lado, y los elementos de la realidad, por otro, pueden combinarse» (Cerezo, 2003: 35).

en todo caso, mantiene cierto paralelismo con los juegos de lenguaje atribuidos al segundo Wittgenstein (2009). Por ejemplo, una silla, en un determinado contexto fenomenológico y cognitivo, puede actualizarse en una mesa, y en este mismo contexto, el suelo puede a su vez actualizarse en una silla. Según lo anterior, no existe una determinación espacio-temporal para los significados que le son atribuidos a las diversas entidades del mundo ya que constantemente se pueden actualizar en función de la complejidad de todo un contexto de sensaciones y emociones, creencias, deseos o aspiraciones influidas por culturas locales y globales en constante cambio, entre otros muchos aspectos que conducen a un sinnúmero de reacciones diversas en circunstancias semejantes.

Como sospechó el tecnólogo y filósofo Pierre Lévy (1999), más allá de la dialéctica virtualizante que, en determinadas circunstancias fenomenológicas y cognitivas permite «ver doble» los objetos del mundo (como cuando el ser humano primitivo logró ver un garrote en la rama —sustitución de un objeto— cuando se le presentaron determinadas circunstancias biológicas y culturales), existe un proceso de virtualización retórica que «se parece a las operaciones de creación del mundo humano, tanto en el ámbito del lenguaje como en el técnico o relacional: invención, composición, estilo, memoria, acción. Brote ontológico en bruto, la creación se sitúa más allá de la utilidad» (Lévy, 1999: 86).

El acto *retórico* propuesto por Lévy (1999) se constituye entonces a partir de una forma de comprensión y entendimiento que presupone la posibilidad de la enacción y la consecuente actualización (semántica) del mundo pues la acción, justificada en la experiencia, «plantea preguntas, dispone tensiones y propone finalidades; las hace entrar en escena, las pone en juego en el proceso vital. La invención suprema es la invención de un problema, la apertura de un vacío en medio de lo real.» (*Ibid*).

La «enacción» de los seres humanos en *su mundo* imposibilita que el «significado» pueda algún día llegar a definirse o estabilizarse. Las declaraciones sobre el mundo o el entorno que constantemente realiza un sujeto-emisor no son fieles a lo que toma otro sujeto-receptor. Por esta razón se puede decir que la manipulación de símbolos que realiza una IA *no* puede considerarse un proceso con arreglo a informar imparcialmente sobre el sentido y significado que mantiene el mundo en un determinado contexto espacio-temporal, ya que, como sugirió Owen Barfield, «una cosa funciona como símbolo cuando no solo anuncia, sino que representa algo distinto de sí misma» (Barfield, 2018: 18).

Según este filósofo inglés, las formas geométricas del mundo exterior no deben ser consideradas únicamente como signos que nos recuerdan lo que dichas

formas son «en sí mismas», sino también, el tipo de organización simbólica que nos permite activar la dinámica de los conceptos a razón de determinados estados psíquicos y fisiológicos fundamentalmente enactivos. Esto, de algún modo, es lo que ha permitido el continuo desarrollo del lenguaje y el pensamiento abstracto que utilizamos los humanos; seres biológicos dotados de experiencia y estados mentales. Por esta razón se puede decir que la «significación simbólica es inherente a las propias formas del mundo exterior. […] Si el lenguaje «está lleno de sentido», también la naturaleza está llena de sentido» (Barfield, 2018: 18).

En este punto, Barfield (2018) parece coincidir plenamente con la filosofía del organismo de Alfred N. Whitehead (2021) quien llegó a considerar al universo como un «huerto de valores». Si el lenguaje, dada la complejidad humana, es incapaz de transmitir «fielmente» la información, esto es porque el mundo del ser humano, más que un mero compendio de datos y símbolos traducibles de manera sintáctica, tiene *sentido*. Esta es la razón por la que «el mundo natural sólo puede ser comprendido en profundidad como una serie de imágenes que simbolizan conceptos» (Barfield, 2018: 19).

En esta misma línea, Ralph W. Emerson, quien fuera citado por Barfield (2018), sabía muy bien que las palabras eran *emblemáticas* porque las cosas mismas lo eran. Si el mundo, tal como se encuentra, no puede nunca llegar determinarse en un solo significado, no es sensato esperar que el lenguaje sí pueda cumplir con esta tarea. Si el lenguaje supone una forma de representación que «depende» de la complejidad simbólica que le es inherente a la experiencia y cognición humana, entonces las capacidades de la IA se vislumbran como una expresión más de la gran eficiencia de la comprensión y del pensamiento simbólico que puede llegar a producir el ser humano en su entorno y en su interacción social.

Parecería que la complejidad simbólica que nació en la inextricable relación entre el ser humano y su entorno fue lo que de algún modo llevó al filósofo Saul A. Kripke (2005) a deducir que los nombres (de los objetos de dicho entorno) solo podían tener referencia pero *no* el «sentido» y significado que le es propio a la complejidad de la experiencia y cognición de la vida humana, es decir, a la extensión fenomenológica en la que se produce el conocimiento *humano* del entorno *humano*.

Si se considera que lo descrito hasta aquí no es el caso y por lo tanto se acepta la posibilidad de que la IA algún día sea capaz de desarrollar ciertas capacidades cognitivas de manera «independiente» al tipo de cognición que genuinamente se deriva de la experiencia humana, entonces, habrá que especular cuidadosamente sobre el *sentido* que la IA le atribuye a su mundo artificial, o, en todo caso, al tipo

de significados que subyacen a la comprensión de un tipo de *Umwelt*[11] o entorno circundante que inevitablemente le debería ser correlativo a su constitución ontológica, de otro modo, habrá que hacer una regresión generalizada a una forma ortodoxa de positivismo que permita asumir, de manera gratuita, que existe un modo de comprensión neutral o libre de valores que puede ser apto para determinar la vida humana en detrimento de la complejidad de cualquier forma de mundo simbólico.

De aceptarse lo anterior, la capacidad (puramente instrumental) de la IA acabará considerándose superior a la capacidad simbólica del ser humano (de hecho, esto último ya ha sido planteado seriamente por algunos filósofos e investigadores de la IA próximos a una forma de eliminativismo[12]), y llegado este punto, el ser humano habrá creado un mundo definitivo, es decir, un mundo en el que se ha fijado el significado en aras de alcanzar la literalidad de su sentido.

En estas circunstancias, el ser humano, como escribió el urbanista e historiador Lewis Mumford hace ya más de medio siglo, habrá

> creado un mundo al revés en el que las máquinas se habían hecho autónomas y los hombres se habían convertido en seres serviles y mecánicos, es decir, condicionados por las cosas, externalizados, deshumanizados, desconectados de sus valores y sus objetivos históricos. (Mumford, 2014: 42)

En resumen, si la IA tuviera la capacidad de pensar y comprender independientemente de la producción de los significados humanos ¿qué sería aquello que estaría comprendiendo? ¿podrán «nuestras» mesas, arboles y edificios significar *algo* para la IA «independientemente» de los significados que nosotros continuamente les asignamos?

Si no es el caso, entonces la IA está incapacitada para decodificar un mundo que sea independiente del mundo simbólico que ha sido construido en la médula de la fenomenología del ser humano, y en este sentido, las capacidades de la IA seguirán siendo «dependientes» de la cognición humana.

---

[11] *Umwelt*. Según el biólogo alemán Jakob von Uexküll, el *Umwelt* no es otra cosa que el mundo circundante que le es propio a un determinado organismo. Esta relación, Uexküll la explica en una conocida frase: «el sol del mundo circundante porta la medida del ojo y el ojo del ser vivo porta la medida del sol de su mundo. Así de distintos son los ojos de los seres vivos como de diferentes los soles y los cielos de sus mundos circundantes» (Uexküll, 2014: 92-93).

[12] Eliminativismo. Materialismo en filosofía de la mente. Lo anterior según la acepción que aparece en Abbagnano, N. (2016): *Diccionario de filosofía,* CdeMx: FCE.

Esto último continuará ocurriendo aún cuando el ser humano haya decidido someterse, por decisión propia, al dictado de las capacidades puramente instrumentales de la IA.[13]

Por el contrario, si se admite que la IA «depende» de nuestra experiencia para pensar o comprender de la manera en que lo hace, entonces la IA no estaría pensando o comprendiendo sino simplemente reaccionando a las entradas de contenido que en todo caso se desarrollan de manera creativa a partir de la experiencia y el entendimiento de un ser humano que inevitablemente se disuelve en un entorno de cambios.

En este escenario, la IA aún estaría conservando su lugar como «extensión instrumental» de lo que Dreyfus (1992) llamó *inteligencia general*.

Discusión

Si bien, como afirma Crane (2022), la argumentación de Searle (1980) no tiene en cuenta que la idea de pensar y comprender implica realizar cierta computación simbólica —además de otros procesos (cognitivos) que le pueden ser atribuidos al pensamiento— es importante tener en cuenta que la supuesta comprensión o entendimiento que se le atribuye a la IA *es un proceso que depende de las representaciones mentales (simbólicas) propias de una inteligencia general como la del ser humano que hace continuidad con su entorno*. Esto puede entenderse de esta manera, si se considera que los símbolos o representaciones con las que interactúa la IA no pueden interpretarse a sí mismos pues la sola «sintaxis no es suficiente para la semántica» (Crane, 2022: 206).

Por otro lado, la razón de que el presente análisis se haya enfocado en los paradigmas simbólico y enactivo de la IA fue porque, desde el punto de vista biológico, existe una relación de interdependencia entre los sistemas simbólico y enactivo en aquellos organismos dotados con *inteligencia general*. No obstante, «la tesis de la dependencia cognitiva», en principio, debería poder aplicarse a otros paradigmas más desarrollados (como por ejemplo el conexionismo de redes neuronales) si se considera que la dinamicidad simbólica (semántica) en su relación

[13] El enfoque instrumental de la IA ha sido claramente descrito por el profesor Alva Noë en un ejemplo: «de la misma manera que un reloj de pulsera no sabe la hora, a pesar de que la usamos para saberla, el ordenador no entiende las operaciones que hacemos con él. Nosotros pensamos con los ordenadores, pero los ordenadores no piensan: son herramientas. Si los ordenadores son procesadores de información lo son de la misma manera que los relojes. Y ese hecho no ayuda a entender los poderes de la cognición humana» (Noë, 2010: 201).

con la fenomenología enactiva suponen procesos imprescindibles para el desarrollo de las redes neuronales (naturales) que la IA intenta emular.

En una palabra, la relación simbólica-enactiva característica de una *inteligencia general* dotada con una compleja fenomenología en la que se implica la coextensividad cerebro-cuerpo-entorno, en cierto modo representa el punto de referencia al que aspira la Inteligencia Artificial Fuerte (IAF). Desde esta perspectiva, la coextensividad de la *inteligencia general* (biológica) no debe comprenderse como un atributo prescindible, sino como el único medio que permite el desarrollo eficiente no solo de la complexión fisiológica de los organismos vivos, sino de la *cualidad* de la experiencia que en última instancia influirá en el tipo particular de decisiones que justificarán sus necesidades adaptativas más profundas.

Conclusión

Si la argumentación aquí presentada es consistente, las consecuencias que se deben considerar con especial cuidado no deberían ser aquellas que concuerdan con la posibilidad de que la IA pueda o no alcanzar la capacidad de comprensión o de pensamiento, sino más bien, con el hecho de que las capacidades de la IA *se admitan* como procesos genuinos de comprensión que puedan considerarse legítimamente independientes. La consecuencia de esto podría ser similar a las implicaciones derivadas del «automatismo» con el que los seres humanos se habitúan a un mundo en el que se ignora la necesidad del cambio y la evolución. En otras palabras, si el ser humano, en su búsqueda de la máxima comodidad reconoce con legitimidad a la IA como una inteligencia independiente y superior a sus propias facultades físicas y mentales, la humanidad (a través de la IA) quedará sometida al dictado de sus hábitos más arraigados. En estas circunstancias, la inteligencia y la creatividad[14] podrían perder el equilibrio (recordar la proporción entre el hábito de las semejanzas y la

[14] En primer lugar, entiéndase aquí «inteligencia» con la acepción utilizada en el ámbito concreto de la IA. En este ámbito, como ha sugerido Crane (2022), la inteligencia artificial (computacional) de alguna manera participa del pensamiento, pero no así del tipo de pensamiento o comprensión mental en el que un ser humano, en continuidad con su entorno, adquiere la facultad de actualizar constantemente el significado de la sintaxis a razón de sus necesidades fundamentalmente biológicas y sociales. Este proceso, por su continua novedad, puede ser considerado «creativo». Por lo tanto, el equilibrio entre inteligencia y creatividad supone considerar las facultades humanas de pensamiento que de algún modo están presentes en la IA junto aquellas otras que de ningún manera podrían estar presentes ahí, tales como los procesos creativos (y adaptativos) que, como se ha sugerido, requieren de una relación *fenomenológica* con un entorno en continua transformación.

creación de las diferencias descritas al principio) que para bien o para mal, ha originado cierto progreso civilizatorio. Por lo tanto y dada la continuidad entre el cerebro, el cuerpo y el entorno, la perdida de este equilibrio inevitablemente tendrá repercusiones en la adaptabilidad del ser humano en su medio.

Si por el contrario la IA es consecuente con el equilibrio entre la inteligencia y la creatividad del ser humano, entonces la IA seguirá siendo una eficiente extensión instrumental alimentada por el mundo simbólico que origina dicha inteligencia y creatividad.

Marcos de J. Aguirre Franco
Universidad de Guadalajara
marcosdej.aguirre@gmail.com

Bibliografía

Arias, A. (2021): *Introducción a la ciencia de la conciencia: el estudio de la experiencia subjetiva en filosofía, psicología y neurociencias*, Madrid, Editorial Catarata.

Barfield, O. (2018): *El arpa y la cámara*, Girona, Atalanta.

Cerezo, M., Acero, J. (ed. lit.), Flores, L. (ed. lit.), y Flórez, A. (ed. lit.) (2003): "Viejos y nuevos pensamientos: ensayos sobre la filosofía de Wittgenstein: Isomorfismo y proyección en el Tractatus", Editorial Comares, pp. 31-51.

Crane, T. (1995): *The Mechanical Mind: A Philosophical Introduction to Minds, Machines, and Mental Representation,* London, Routledge.

Crane, T. (2022): *La mente mecánica: Introducción filosófica a mentes, máquinas y representación mental*, CdeMx: FCE.

Dreyfus, H. (1992): *What Computers Still Can't Do: A critique of artificial reason*, Cambridge, The MIT Press.

Gausa, M., Guallart, V., Müller, W., Soriano, F., Porras, F., y Morales, J. (2000): *Diccionario metápolis de arquitectura avanzada: ciudad y tecnología en la sociedad de la información*, Barcelona, Editorial Actar.

Kripke, S. (1981): *Naming and Necessity*, Oxford, Basil Blackwell Publisher.

Lévy, P. (1999): *¿Qué es lo virtual?* Buenos Aires, Paidós Multimedia.

Mumford, L (2014): *Arte y técnica,* Logroño, Editorial Pepitas de Calabaza.

Noë, A. (2010): *Fuera de la cabeza: Porqué no somos el cerebro. Y otras lecciones de la biología de la consciencia*, Barcelona, Kairós.

Searle, J. (1980): "Mind, brains and programs: Behavioral and brain science", Issue 3, pp. 417-424.

Uexküll, J. (2014): *Cartas biológicas a una dama*, Buenos Aires, Editorial Cactus.

Whitehead, A. N. (2021): *Proceso y realidad*, Girona, Atalanta.

Wittgenstein, L. (2009): *Philosophical Investigations,* New Jersey, Wiley-Blackwell.