El concepto de inteligencia: el *hard problem* de la inteligencia artificial*

The Concept of Intelligence: the *Hard Problem* of Artificial Intelligence

Oscar D. Caicedo†
Eduardo Bermúdez Barrera‡

Resumen

En el amplio marco interdisciplinario que conforma el hexágono cognitivo —filosofía, antropología, psicología, lingüística y neurociencia—, la Inteligencia artificial se ha posicionado probablemente como el ámbito que más controversias y desacuerdos ha suscitado en los últimos años. En este artículo se defiende que el concepto de inteligencia como normalmente se aplica a la IA es demasiado generoso, creando confusiones en torno a lo que tradicionalmente —desde las ciencias cognitivas y la biología, por ejemplo—, se entiende por pensamiento o comportamiento inteligente. Contrario a la idea popular de que la inteligencia consiste llanamente en resolver problemas, proponemos que la inteligencia es más que eso: es resolver problemas de manera flexible y situacional, haciendo predicciones a corto y largo plazo de manera automática y que ser asertivos en dichas predicciones, muchas veces, es lo que posibilita nuestra supervivencia.

Palabras clave: cognición - resolución de problemas - memoria - predictibilidad

Abstract

Within the entire interdisciplinary framework that makes up the cognitive hexagon—philosophy, anthropology, psychology, linguistics and neuroscience—, Artificial Intelligence is probably the field that has generated most debate and disagreement in recent years. This article argues that the concept of intelligence as it is usually applied to AI is too generous, creating confusion around what is traditionally understood—from neuroscience and biology, for example—as intelligent thinking or behavior. Contrary to the popular idea that intelligence is about solving problems, we propose that intelligence is more than that: it is about solving problems in a flexible and situational way, making short- and long-term predictions automatically, and that being assertive in these predictions is often what makes our survival possible.

Keywords: cognition - problem solving - memory - predictability

* Recibido: 2 de septiembre de 2025. Aceptado con revisiones: 30 de septiembre de 2025.
† Universidad del Atlántico, Colombia. Grupo de Investigación Holosapiens. Para contactar al autor, por favor, escribir a: oscarcaicedo@mail.uniatlantico.edu.co
‡ Universidad del Atlántico, Colombia. Grupo de Investigación Holosapiens. Para contactar al autor, por favor, escribir a: eduardobermudez@mail.uniatlantico.edu.co
Metatheoria 16(1)(2025):): 1-12. ISSN 1853-2322. eISSN 1853-2330.

1. Sobre el concepto de inteligencia

Ofrecer una definición de inteligencia no es una tarea fácil y parte de la dificultad radica en que los términos de la definición varían en función de la disciplina que la ofrezca (Colmenares, 2015). Desde la biología —y la filosofía de la biología—, p. ej., se suele decir que un individuo (o especie) tiene un comportamiento inteligente cuando recurre a procesos cognitivos complejos para sortear los problemas que le plantea el medio, tanto físico como social.

En ese sentido, escribe Richard Byrne:

> En los humanos, la inteligencia ciertamente significa más que aprendizaje flexible: términos como "pensar claramente", "resolver problemas difíciles" y "razonar bien" recurren a los intentos de definir esta capacidad. El alcance de la inteligencia es bastante amplio, incluyendo aprender una gama de información sin restricciones, aplicar este conocimiento en otras y quizás nuevas situaciones, beneficiarse de las habilidades de los demás y pensar, razonar o planificar nuevas tácticas, lo que debería recordarnos que no debemos esperar que la inteligencia sea una sola cosa, sino un conjunto de dispositivos y procesos, dotaciones y aptitudes que juntos producen un comportamiento que vemos como "inteligente" (Byrne 1995, p. 38).

La definición de Byrne supone así que la inteligencia permite al individuo aumentar su conocimiento a través de interacciones con el mundo físico y social, utilizar ese conocimiento para modular su comportamiento —tanto en contextos conocidos como poco familiares— y relacionar e integrar elementos de conocimiento inicialmente inconexos que le permitan crear acciones nuevas por medio del razonamiento y la planificación.

La inteligencia no es entonces un vaciado de información en un recipiente (el cerebro), tener la habilidad de resolver operaciones matemáticas con grandes cifras, recordar en detalle cada palabra de un libro leído o ganar partidas de ajedrez de gran complejidad. La inteligencia tampoco se obtiene súbitamente con un *click*, sino que es un proceso lento y paulatino que se va desarrollando y depurando a medida que interactuamos con otros y con el mundo físico mientras nos desarrollamos ontogenéticamente.

Sobre si es pertinente o no llamar inteligente a la IA, escribe Carlos Madrid Casado que "el *quid* de la cuestión es que la IA no es, en puridad, inteligencia ni artificial, si por inteligencia se entiende algo parecido a la humana y si por artificial se entiende algo contrapuesto a natural (como suponen que es la inteligencia humana)" (Madrid Casado 2024, p. 87). He considerado que argumentar que la IA es inteligente porque, por ejemplo, realiza tareas complejas o puede "aprender" de la experiencia puede ser demasiado flexible. El comportamiento inteligente que mejor conocemos es el nuestro; también conocemos —aunque con algunas limitaciones— el comportamiento inteligente en otras especies animales. Del primero podemos sostener que está provisto de conciencia, creatividad (en el sentido estricto del término), la capacidad de razonamiento abstracto, voluntad e intuición. *Sentir el mundo* hace parte del proceso automático (cognitivo-perceptivo) de aprehender la realidad, movernos en ella y tomar decisiones. Esa toma de decisiones y ese moverse en el mundo no están necesariamente atados a algoritmos y modelos matemáticos para procesar y analizar datos. Difícilmente podríamos hablar en los mismos términos -al menos no sin muchas reservas- de cualquier IA conocida hasta ahora.

Sostiene Erik Larson:

> No existe ningún algoritmo para la inteligencia general. Y tenemos buenos motivos para mostrarnos escépticos ante la idea de que dicho algoritmo vaya a surgir de nuevas tentativas con los sistemas de aprendizaje profundo o de cualquier otra aproximación popular en la actualidad. Resulta mucho más probable que vaya a requerir de un avance científico de primer orden, y ahora mismo nadie tiene la más remota idea del aspecto que tendría ese avance, y mucho menos de los detalles que conducirán a él (Larson 2022, p. 8).

Sugiere Larson que la mayor muestra de genio por parte de Turing, y también su mayor error, consistió en pensar que la inteligencia humana se limitaba a resolver problemas, pero, además, en la creencia de que existía algo así como una inteligencia general, completamente abstracta, desconectada de su contexto.

Por su parte, el desarrollador de software francés François Chollet ha escrito que la inteligencia es *situacional* en estos términos:

> El primer problema que veo con la teoría de la explosión de la inteligencia es que no reconoce que la inteligencia es necesariamente parte de un sistema más amplio, una visión de la inteligencia como un "cerebro en un frasco" que puede volverse arbitrariamente inteligente independientemente de su situación. Un cerebro es solo un trozo de tejido biológico, no tiene nada de inteligente intrínsecamente. Más allá del cerebro, el cuerpo y los sentidos (las capacidades sensoriomotoras) son una parte fundamental de la mente. El entorno es una parte fundamental de la mente. La cultura humana es una parte fundamental de la mente. Al fin y al cabo, es de ahí de donde provienen todos nuestros pensamientos. No se puede disociar la inteligencia del contexto en el que se expresa [...] no existe una inteligencia "general". [...] Si la inteligencia es un algoritmo de resolución de problemas, entonces solo puede entenderse con respecto a un problema *específico*. De una manera más concreta, podemos observar esto empíricamente en el hecho de que todos los sistemas inteligentes que conocemos están altamente especializados. La inteligencia de las IA que construimos hoy está hiperespecializada en tareas extremadamente limitadas (Chollet 2017).

Concluye Chollet que (1) el cerebro es una pieza de un sistema más amplio que incluye el cuerpo, el entorno, otros seres humanos y la cultura, (2) toda inteligencia individual siempre estará definida y limitada por el contexto de su existencia, por su entorno. Nuestro entorno, no nuestro cerebro, es el que actúa como un cuello de botella para nuestra inteligencia, (3) la inteligencia humana no está contenida en nuestro cerebro sino en nuestra civilización. Somos nuestras herramientas: nuestros cerebros son módulos de un sistema cognitivo mucho más grande que nosotros mismos. Un sistema que ya se está perfeccionando a sí mismo, y lo ha estado haciendo durante mucho tiempo, (4) la ciencia es posiblemente el sistema más cercano a una IA que se mejora recursivamente a sí misma que podemos observar y (5) la expansión de la inteligencia recursiva ya está ocurriendo, a nivel de nuestra civilización. Seguirá ocurriendo en la era de la IA y progresa a un ritmo aproximadamente lineal.

Aunque Turing creyó que la intuición, por ejemplo, se puede programar en una máquina,[1] algunos autores —Chollet (2017) entre ellos— sostienen que esta no podrá alcanzar el nivel de la inteligencia humana. Los seres humanos disponemos de inteligencia social y emocional y nuestras mentes no solo resuelven problemas y acertijos (Larson 2022).

Kate Crawford (2024) abre su *Atlas de inteligencia artificial* con la curiosa historia de "Clever Hans", un caballo alemán que a finales del siglo XIX deslumbró a muchos por su aparente capacidad para, por medio de golpeteos con el casco, dar solución a algunas operaciones matemáticas y responder acertadamente cuando se le preguntaba por los días de la semana. Luego de algunas pruebas exhaustivas que incluyó a académicos, zoólogos, veterinarios y oficiales de caballería, se concluyó que, acciones como la postura del interrogador, su respiración y expresión facial, por ejemplo, cambiaban cuando Hans alcanzaba la respuesta correcta por medio de los golpeteos de su casco, haciendo que al caballo cesara de golpear justo en ese preciso momento. Lo curioso es que, según los investigadores, los interrogadores daban las pistas al caballo de manera inconsciente e involuntaria.

De lo sucedido con Hans, Crawford describe dos mitologías distintas que vale la pena considerar:

[1] Aquí el término "máquina" se emplea en un sentido restringido, referido a sistemas artificiales computacionales. Se dejan de lado los sistemas biohíbridos o biomáquinas que integran componentes biológicos, ya que, aunque estos complejizan la distinción entre lo natural y lo artificial, no afectan el núcleo del argumento del artículo, centrado en las condiciones funcionales, contextuales y situacionales de la inteligencia.

> El primer mito es que los sistemas no humanos (sean computadoras o caballos) son análogos a la mente humana. Esta perspectiva presupone que, con el entrenamiento adecuado o los recursos suficientes, una inteligencia parecida a la del ser humano se puede crear de cero sin tener en consideración las maneras fundamentales en que las personas se encarnan, se relacionan y se ubican dentro de contextos más amplios. El segundo mito es que la inteligencia es algo que existe de forma independiente, como algo natural y separado de las fuerzas sociales, culturales, históricas y políticas (Crawford 2024, pp. 23-24).

Estos dos mitos, según la autora, prevalecen en el campo de la IA, campo en el que las personas creen seriamente que las máquinas pueden reproducir la inteligencia humana. Por otro lado, creer que los seres humanos somos máquinas que resuelven problemas o sistemas procesadores de información lleva al riesgo de sugerir una idea muy simple de lo que es la inteligencia.

Pero no todos han sido tan optimistas. Durante las conferencias "Administración de empresas y la computadora del futuro", realizadas en el MIT en 1961 con la participación de Allen Newell, Herbert Simon, Norbert Wiener, entre otros, el filósofo Hubert Dreyfus sugirió que el cerebro procesaba información de manera diferente a una computadora. Algunos años más tarde, en *What Computers Can't Do*, Dreyfus (1972) sostiene que la inteligencia humana depende, en gran medida, de procesos inconscientes y subconscientes, mientras que las computadoras requieren que todos sus procesos y datos estén formalizados.

En el capítulo 10, que Dreyfus tituló sugerentemente "The Limits of Artificial Intelligence", el autor se pregunta: (1) ¿un ser humano en el "procesamiento de información" realmente sigue reglas formales como una computadora digital?, y (2) ¿puede el comportamiento humano, sin importar cómo se genere, describirse en un formalismo que pueda ser manipulado por una máquina digital?, a lo que responde:

> la evidencia descriptiva o fenomenológica, considerada aparte de los prejuicios filosóficos tradicionales, sugiere que capacidades humanas no programables están involucradas en todas las formas de comportamiento inteligente. [...] en la medida en que la cuestión de si la inteligencia artificial es posible es una cuestión empírica, la respuesta parece ser que es extremadamente improbable que se produzcan nuevos avances significativos en la simulación cognitiva o en la inteligencia artificial (Dreyfus 1972 p. 197).

Hay que decir, sin embargo, que desde entonces los avances y cambios en las investigaciones en torno a la inteligencia artificial han sido inmensos y gran parte de los primeros debates han quedado en el olvido. La IA hoy no es solo un campo de investigación, sino una industria que genera mucho dinero, aunque persiste aún la pregunta antropomórfica de si puede o podrá una máquina emular la cognición humana.

Crawford sostiene categóricamente que "la IA no es *artificial* ni *inteligente*":

> Más bien existe de forma corpórea, como algo material, hecho de recursos naturales, combustible, mano de obra, infraestructuras, logística, historias y clasificaciones. Los sistemas de IA no son autónomos, racionales ni capaces de discernir algo sin un entrenamiento extenso y computacionalmente intensivo, con enormes conjuntos de datos o reglas y recompensas predefinidas (Crawford 2024, p. 29).

En el mismo sentido, el científico de la computación Luc Julia (2019) ha afirmado que la inteligencia artificial no existe, no al menos en el sentido en que mucha gente la ha entendido o en los términos rodeados de promesas fabulosas alrededor del cual se ha tejido la idea.

Es cierto que *Deep Blue* derrotó a Kasparov en ajedrez en 1997, ¿pero era Deep Blue más inteligente que éste? En términos estrictos no se trata de inteligencia. Deep Blue aprendió las reglas del ajedrez y se puso en su memoria tantas partidas como fuera posible, para que la máquina pueda encontrar la mejor jugada en su base de datos. *Es un proceso de fuerza bruta*, sostiene Julia. La máquina ganó porque tenía la capacidad, gracias a su memoria, de predecir más movimientos con más antelación que Kasparov.

Otro tanto ocurrió con el juego del Go. Dos decenios después de *Deep Blue*, *AlphaGo* (un programa de *Google DeepMind*) derrotaba al campeón mundial Lee Se-dol, en un duelo de cinco partidos. Aquí la estrategia es diferente a la utilizada en 1997: en lugar de darle al programa una biblioteca de

movimientos, lo entrenaron para jugar a partir de un conjunto de 30.000 partidas. Este conjunto es muy pequeño en comparación con la cantidad de partidas posibles, sin embargo, AlphaGo, gracias a estas partidas y su análisis, ha adquirido una habilidad que le ha permitido, no solo vencer al campeón del mundo, sino también ofrecer jugadas sorprendentes que, según varios grandes jugadores, han cambiado su propia concepción del juego. Pero para hacer que una máquina venza a un humano en el Go —y sólo eso, porque es hiperespecializada— hay que gastar grandes cantidades de energía (un verdadero centro de datos que consume 440 kWh) donde un humano consume solo unos pocos vatios (Julia 2019).

La inteligencia artificial no es inteligencia, sostiene Julia. Tampoco es una caja negra que escapa a nuestro poder. Esto es algo sobre lo que tenemos control total. Si se producen errores, se debe a errores en los algoritmos o en los datos. Y estos errores siempre se pueden explicar.

> Todo es explicable, en IA, en teoría. En la práctica, las cosas pueden ser más complicadas. Estamos hablando de máquinas que hacen unos pocos millones de cálculos por segundo. Si tenemos que volver a pasar por todos los cálculos que ha hecho la máquina, también nos llevará unos millones de segundos, que no tenemos. Por lo tanto, podemos hablar de inexplicable, pero solo por razones prácticas. Matemáticamente, todo lo que llamamos inteligencia artificial es totalmente explicable (Julia 2019, pp. 12-13).

Actualmente es posible crear algoritmos que son capaces, a partir de miles de imágenes de un objeto que hay en la red, de reconocer ese objeto con un altísimo porcentaje de precisión. La inteligencia humana no necesita esos miles de imágenes previas de ese objeto, es multimodal: utiliza no solo lo que percibimos visualmente, sino también los otros sentidos, el contexto y muchas otras cosas que son mucho más difíciles de modelar que una imagen. La inteligencia humana sabe inventar: es capaz de adaptarse a nuevas situaciones e inventar sin haber sido específicamente preparada para una situación en particular. La respuesta de la inteligencia humana a los desafíos del medio está influenciada por la cultura, influencia que en parte es inconsciente. Aunque una máquina puede perseguir un objetivo establecido por un programador, dice Julia, no puede hacerlo de *motu proprio*.

2. La predicción es la base de la inteligencia

Con Jeff Hawkins (2023, Hawkins & Blakeslee 2004), creador de Palm Pilot, encontramos la otra cara de la moneda. Hawkins sostiene que si bien es complejo definir la inteligencia ("la última gran frontera terrestre de la ciencia") sí podemos decir en qué consiste ser inteligente: predecir para resolver problemas y esa predicción se hace con patrones incompletos almacenados en la memoria. "Es en la capacidad de hacer predicciones sobre el futuro donde está el *quid* de la inteligencia" (Hawkins & Blakeslee 2004, p. 6).

Aunque hay hoy miles de neurocientíficos en todo el mundo suministrando una gran cantidad de datos sobre el cerebro y su funcionamiento, no habrá teorías productivas —dice Hawkins— hasta que piensen más en teorías generales del cerebro y menos en hacer experimentos para recopilar más datos sobre los subsistemas de éste. Una de las tesis principales del autor es que es necesario comprender cómo funciona el neocórtex con suficiente detalle para poder explicar la biología del cerebro y construir máquinas inteligentes que se basen en los mismos principios (Hawkins 2023).

En una posición bastante optimista que contrasta con la de Larson, sostienen Hawkins y Blakeslee:

> Podremos construir máquinas genuinamente inteligentes, aunque no se parecerán en nada a los robots de la ficción popular y la fantasía informática. Más bien, las máquinas inteligentes surgirán de un nuevo conjunto de principios sobre la naturaleza de la inteligencia. Como tales, nos ayudarán a acelerar nuestro conocimiento del mundo, nos ayudarán a explorar el universo y harán que el mundo sea más seguro. Y en el camino, se creará una gran industria (Hawkins & Blakeslee 2004, p. 2).

Como se mencionó, Hawkins sostiene que la capacidad para hacer predicciones está directamente relacionada con la inteligencia. Defiende que nuestro cerebro emplea memorias almacenadas para realizar predicciones constantes sobre todo lo que vemos, sentimos y escuchamos. Cuando en este preciso momento miro alrededor de mi biblioteca, mi cerebro está utilizando recuerdos para formar predicciones sobre lo que espera encontrar antes de que suceda. La gran mayoría de las predicciones —dice— ocurren sin que tengamos conciencia de ello, pero cuando entra un patrón visual que no había memorizado en ese contexto (p. ej. una taza de café desconocida para mí), la predicción se incumple y mi atención se dirige al error.

> Nuestro cerebro efectúa predicciones constantes sobre la misma estructura del mundo en que vivimos, y lo hace en paralelo. Estará igual de dispuesto para detectar una textura extraña, una nariz deforme o un movimiento inusual. El carácter omnipresente de estas predicciones, inconscientes en su mayoría, no se advierte a primera vista, motivo por el cual tal vez hemos pasado por alto su importancia durante tanto tiempo. Suceden de modo tan automático, con tanta facilidad, que no logramos desentrañar lo que está pasando dentro de nuestros cráneos (Hawkins & Blakeslee 2004, p. 87).

Pero la predicción no se limita a patrones de información sensorial de bajo nivel como ver y escuchar. El cerebro humano puede hacer predicciones sobre tipos de patrones abstractos y secuencias de patrones temporales largas. La propuesta de Hawkins es que la inteligencia superior no es un tipo diferente de proceso de la inteligencia perceptiva. Se basa en lo fundamental en la misma memoria de la corteza cerebral y algoritmo de predicción.

> La inteligencia se mide por la capacidad de recordar y predecir patrones del mundo, incluidos lenguaje, matemática, propiedades físicas de los objetos y situaciones sociales. Nuestro cerebro percibe patrones del mundo exterior, los almacena como memoria y realiza predicciones basadas en las comparaciones entre lo que ha visto antes y lo que sucede ahora (Hawkins & Blakeslee 2004, p. 97).

La piedra de toque de la inteligencia (al menos de la humana y la de otros mamíferos) es la corteza cerebral. La memoria y la predicción permiten a un animal utilizar sus conductas existentes (el cerebro viejo) de forma más inteligente. Para realizar predicciones sobre acontecimientos futuros, nuestra corteza cerebral tiene que almacenar secuencias de patrones. Para recordar las memorias apropiadas tiene que recuperar patrones por su similitud a patrones pasados (recuerdo autoasociativo). Y, por último, las memorias han de guardarse en una forma invariable a fin de que el conocimiento de acontecimientos pasados sea aplicable a nuevas situaciones que son similares, pero no necesariamente idénticas al pasado (Hawkins & Blakeslee 2004).

Pero ¿cuál es el papel de la memoria y la predicción en la inteligencia como la conocemos? La competencia más básica a la que se enfrentan todos los animales es entre depredadores y presas. Todos necesitamos reconocer, muy rápidamente, si nos enfrentamos a un amigo o a un enemigo. Desde el nacimiento, la mayoría de los animales reconocen rápidamente a otros agentes a partir de sus movimientos. Los agentes no son como los objetos físicos. Son autopropulsados, aunque sus movimientos están limitados por su estructura física. A corto plazo, estas limitaciones permiten predecir sus movimientos. A largo plazo, los movimientos se pueden predecir en función de sus objetivos. Los agentes dirigidos a objetivos se comportan racionalmente, en el sentido de que evitan los obstáculos y alcanzan sus objetivos por el camino más directo disponible. Los bebés humanos en su primer año de vida esperan que los agentes autopropulsados tengan metas y se comporten racionalmente para alcanzarlas. Los bebés también reconocen que las preferencias determinan las metas y que otras personas pueden tener preferencias diferentes. Pero hay un nivel aún más alto en el que podemos reconocer a los agentes como comportándose de acuerdo con sus intenciones internas. Es difícil detectar las intenciones solo a partir de los movimientos. Críticamente, interpretaremos los mismos movimientos de manera diferente si creemos que estamos tratando con un agente intencional. Sin esta suposición, el comportamiento intencional parece irracional. Esto sucede, por ejemplo, cuando las personas exageran

sus acciones para señalar sus intenciones. Este alejamiento del comportamiento de menor esfuerzo, racional y dirigido a un objetivo es una señal segura de que estamos tratando con un agente intencional (Frith & Frith 2023).

Somos una especie intensamente social, profundamente dependientes los unos de los otros para nuestra propia supervivencia, pero también somos una especie complicada, con un repertorio de acciones más amplio que cualquier otra. Tenemos entonces que ser capaces de predecir lo que otras personas harán y una de las mejores maneras de hacerlo es saber lo que está pasando por sus mentes (Gopnik, Meltzoff & Kuhl 1999). Para detectar intenciones en otros, generalmente hace falta percibir su movimiento. El movimiento resulta ser fundamental para detectar agentes, y cualquier movimiento en el mundo exterior atrae nuestra atención y se clasifica instantáneamente. En el cerebro humano, una región circunscrita en el surco temporal posterior superior (pSTS) está sintonizada para detectar agentes en movimiento.

La región pSTS/TPJ parece ser la puerta de entrada desde el mundo físico hacia el sistema de mentalización. En particular, esta región se ocupa de la pregunta: "¿La persona con la que estoy interactuando está haciendo lo que esperaba?". El pSTS es una región que se activa por el movimiento biológico (Grossman *et al.* 2000) y tiene un papel importante en la observación de la acción. Esta área es adyacente a la unión temporo-parietal (TPJ), que ha sido implicada repetidamente en los estudios de mentalización. Varios estudios sugieren que el pSTS/TPJ está implicado en la predicción. Y este es el caso cuando estamos monitoreando nuestras propias acciones, así como las acciones de otras personas. Cuando el resultado de nuestra acción no es el esperado, se observa una mayor actividad en la TPJ (Miele *et al.*, 2011). Cuando estamos monitoreando a otros, se observa una mayor actividad cuando las personas mueven sus ojos en una dirección inesperada (lejos de, en lugar de hacia un estímulo objetivo) (Pelphrey *et al.* 2003).

Sostiene Rodolfo Llinás:

> La predicción de eventos futuros —vital para moverse eficientemente— es, sin duda, la función cerebral fundamental y más común [...] de ella depende la vida misma del organismo. [...] La capacidad del cerebro para predecir no se genera sólo a nivel consciente, ya que evolutivamente la predicción es una función mucho más antigua que la conciencia (Llinás 2001, p. 21).

La idea de que pSTS/TPJ desempeña un papel específico en la predicción también está respaldada por un estudio de Bardi, Gheza y Brass (2017). En este caso, se utilizó la estimulación magnética transcraneal (TMS por sus siglas en inglés) para interrumpir la TPJ (en el lado derecho del cerebro), mientras que los participantes vieron videos que involucraban a un agente con una creencia sobre la ubicación de un objeto. La realización de esta tarea requiere una representación de la creencia del agente (verdadera o falsa) y una predicción de lo que va a suceder. La TMS no interrumpió la representación de la creencia del agente, pero sí interfirió con las predicciones.

Hacer predicciones momento a momento sobre los movimientos futuros puede basarse en los principios físicos asociados con los objetos, como las leyes del movimiento y las restricciones impuestas por el cuerpo. Sin embargo, para hacer predicciones a largo plazo, es útil tener en cuenta los objetivos e intenciones de nuestros oponentes. La mayoría de los animales pueden reconocer agentes dirigidos a un objetivo a partir de su comportamiento racional (por ejemplo, evitar obstáculos), así como de sus preferencias constantes. Los humanos y algunas otras especies pueden hacer predicciones a largo plazo aún mejores sobre el comportamiento de los demás si hacen inferencias sobre sus intenciones y creencias. Pero para predecir con éxito momento a momento hacia dónde irá un agente a continuación, no es necesario tener en cuenta estados ocultos como los objetivos y las intenciones. A corto plazo, las restricciones biológicas sobre los posibles movimientos determinan hacia dónde es probable que se mueva el agente a continuación. Pero a largo plazo, tener en cuenta los objetivos y las intenciones dará una ventaja (Frith & Frith 2023).

Una de las tareas del desarrollo cognitivo de los niños más ampliamente estudiada es la comprensión que tienen sobre la aparición y desaparición de los objetos. Existen cuatro factores que explican cómo los objetos permanentes que se mueven en el espacio llevan a secuencias de aparición y desaparición de objetos. Están, en primer lugar, las leyes que rigen el movimiento de esos objetos, en segundo lugar, están las propiedades de los objetos en sí, en tercer lugar, las relaciones espaciales entre los objetos estáticos y, por último, existen las leyes que gobiernan las relaciones perceptivas entre observadores y objetos (Gopnik & Meltzoff 1999). Nuestro cerebro, a diferencia de cualquier sistema de inteligencia artificial conocido, viene con un *kit* de herramientas que le permite desarrollar estos cuatro factores.

Basados en estos cuatro factores, hacemos predicciones más o menos fiables de cuándo un objeto aparecerá o desaparecerá, pero también cuando no lo hará y cuándo será visible o invisible a un observador concreto. Un buen mago, por ejemplo, crea la ilusión de manejar a su antojo los cuatro factores mencionados y genera desconcierto cuando saca una paloma del bolsillo o cuando desaparece su sombrero sorpresivamente. Nuestra predicción falla.

2.1. El *hard problem* de la inteligencia artificial

En su reciente libro *Mil cerebros*, Hawkins argumenta que,

> el futuro de la IA se basará en principios diferentes de los que se aplican en la actualidad, principios que se asemejarán más al cerebro. Para construir máquinas verdaderamente inteligentes, debemos diseñarlas según los principios basados en el funcionamiento [de éste] [...] la principal razón por la que los sistemas de IA actuales no se consideran inteligentes es que solo pueden hacer una cosa, mientras que las personas pueden hacer muchas. En otras palabras, los sistemas de IA no son flexibles. Cualquier ser humano individual, como tú o yo, puede aprender a jugar al Go, a cultivar, a escribir programas informáticos, a pilotar un avión y a tocar música. Aprendemos miles de habilidades a lo largo de nuestra vida, y aunque no seamos los mejores en ninguna, somos flexibles en lo que podemos aprender. Los sistemas de aprendizaje profundo apenas exhiben alguna flexibilidad. Un ordenador que juega al Go puede jugar mejor que cualquier ser humano, pero no puede hacer nada más. Un automóvil autónomo puede ser un conductor más seguro que cualquier conductor humano, pero no puede jugar al Go ni arreglar un pinchazo (Hawkins 2023, pp. 145-148).

En la pretensión de crear máquinas inteligentes, las investigaciones en el campo han seguido dos caminos principales: por un lado, que de hecho es la tendencia, está el intento —bastante exitoso, de hecho— de crear computadoras que superen a los humanos en tareas específicas, por ejemplo, jugar ajedrez. El otro camino para crear máquinas inteligentes es la apuesta por la flexibilidad: aquí no importa que la máquina haga ciertas tareas mejor que los humanos, sino que sean "capaces de hacer muchas cosas y aplicar lo que aprenden de una tarea a otra" (Hawkins 2023 p. 150).

Ahora la lupa está en lo que se conoce como el problema de la *representación del conocimiento*, lo que llamamos el "*hard problem* de la inteligencia artificial".[2] Mientras un niño de tan solo 5 años adquiere con facilidad gran cantidad de conocimiento cotidiano y conoce y aprende sobre las regularidades y complejidades del mundo diariamente por observación, a los investigadores de IA les ha costado programar este conocimiento cotidiano en una máquina.

> Las actuales redes de aprendizaje profundo no poseen conocimiento. Una computadora que juega al Go no sabe que el Go es un juego. No conoce su historia. No sabe si está jugando contra otra computadora o contra un ser humano, ni qué significan las palabras "computadora" y "humano". De manera similar, una red de aprendizaje profundo que etiqueta imágenes puede mirar una imagen y decir que es un gato. Pero tiene un conocimiento muy limitado de los gatos. No sabe que son animales, ni que tienen cola, patas y pulmones. No sabe nada de los amantes de los gatos frente a los amantes de los perros, ni que los

[2] La expresión "*hard problem*" fue introducida originalmente por David Chalmers para referirse al problema de la conciencia; aquí se utiliza de manera analógica para señalar una dificultad estructuralmente profunda en la representación del conocimiento en sistemas artificiales.

> gatos ronronean y mudan el pelo. Todo lo que hace la red de aprendizaje profundo es determinar que una nueva imagen es similar a las imágenes vistas anteriormente que se etiquetaron como "gato". No hay conocimiento sobre gatos en la red de aprendizaje profundo (Hawkins, 2023 pp. 151-152).

Hawkins no cree que las redes de aprendizaje profundo (*deep learning*) cumplan el objetivo de la inteligencia artificial general a menos que modele el mundo tal como lo hace un cerebro. Aunque las redes de aprendizaje profundo funcionan bien, no han resuelto el problema de la representación del conocimiento, sino que utilizan estadísticas y montones de datos: "las máquinas verdaderamente inteligentes aprenderán modelos del mundo valiéndose de marcos de referencia similares a mapas, como hace el neocórtex. Esto me parece inevitable. No creo que haya otra manera de crear máquinas auténticamente inteligentes" (Hawkins 2023, p. 155).

Indudablemente el tema de las redes neuronales artificiales está en el centro del debate. A la pregunta "¿cómo aprende un programa informático basado en estructuras neuronales?" (Sigman & Bilinkis, 2024, p. 24), los autores sugieren que el mecanismo de conexiones neuronales da origen a un vasto repertorio de aprendizajes que son el cimiento de la asombrosa complejidad de la inteligencia.

> Las redes neuronales artificiales se organizan en una estructura jerárquica de capas sucesivas, otra idea tomada del cerebro humano. En su versión más simple incluyen tres capas: una que codifica la entrada, otra intermedia que la procesa y representa de manera más abstracta, y una de salida para dar una respuesta. Gracias al aumento del poder de cómputo del hardware, fue posible agregar cada vez más cantidad de capas intermedias, dando lugar a un nuevo tipo de red neuronal conocida como aprendizaje profundo [...]. Las redes neuronales combinaron la forma de aprender de las máquinas: ya no se programan con una serie de instrucciones escritas por un humano, sino que se entrenan para que vayan descubriendo los patrones de conexiones neuronales que las vuelven efectivas. En este proceso aparece un elemento que también está en la esencia del aprendizaje humano: la retroalimentación o *feedback* (Sigman & Bilinkis 2024, pp. 25-26).

Algunos autores (p. ej., Rubio, Giri & Ilcic 2023) sugieren que, aunque los modelos de redes neuronales artificiales son una herramienta útil para resolver problemas computacionales complejos, su opacidad epistémica es un tema de discusión, no solo en el campo de la neurociencia sino también en la filosofía de la ciencia. Aun cuando los constructores del modelo conocen muchas características de su arquitectura y funcionamiento interno, dicen los autores, la forma exacta en que una red arroja sus predicciones o clasificaciones generalmente es poco clara. El tránsito entre la entrada y la salida se cubre en la etapa intermedia de procesamiento, dando origen a una especie de "caja negra".

3. No, el cerebro no funciona como una computadora

Roberto Perazzo (1998) definió la inteligencia artificial como la "ciencia"[3] que estudia las facultades mentales por medio de modelos computacionales" (p. 102). Pues bien, indiscutiblemente, las redes neuronales artificiales mencionadas en párrafos anteriores sugieren, al menos en principio, una "adaptación artificial" de las redes neuronales biológicas.

Pero, ¿cómo funcionan las redes neuronales biológicas?

[3] Madrid Casado (2024) ha sugerido -desde la perspectiva de la teoría del cierre categorial- que la IA no es una ciencia sino una pluralidad de tecnologías. Estas tecnologías están constituidas por múltiples técnicas derivadas de diversas ciencias que convergen en el ámbito de la IA —como las matemáticas, la estadística, la física, la biología, la psicología, la neurociencia, la lingüística, entre otras. En este sentido, no se trata de un cierre categorial en sentido estricto (es decir, de naturaleza científica), sino de un cierre fenoménico (de carácter tecnológico), lo que implica que se trata de un ámbito abierto y cambiante. Esto se debe a que las operaciones en el campo de la IA no se articulan de manera interna, estableciendo relaciones autónomas entre términos al margen de los sujetos (situación *alfa*), sino de forma externa (situación *beta*), en tanto que las operaciones se enlazan orientándose todas hacia la consecución de un fin común (p. 34).

> Esta es la idea esencial de una red neuronal. Una malla lo más amplia posible, formada por distintas capas de neuronas idénticas. La combinatoria es tan grande que permite establecer circuitos capaces de codificar casi cualquier cosa. Cada patrón de activación de la red, es decir, cada conjunto de neuronas que se activan de manera simultánea, establece una representación "mental" de un objeto. Puede ser la representación de algo concreto como un animal o de un ente abstracto. Estas estructuras a su vez pueden combinarse para formar representaciones más complejas [...]. Una red neuronal establece una relación unívoca entre los objetos y sus representaciones en grupos específicos de neuronas. Las neuronas que se activan cuando la red ve este objeto se conectan entre ellas. Y en este entramado particular queda el recuerdo de un objeto que puede activarse y representarse de manera abstracta (Sigman & Bilinkis 2024, pp. 24-25).

Ya en 1986, en su artículo "Learning Representations by Back-propagating Errors", Rumelhart, Hinton y Williams (1986) y su equipo propusieron un nuevo procedimiento de aprendizaje, la retropropagación, para redes de unidades similares a neuronas, donde "la dinámica de activación en las neuronas artificiales simularía (imitaría rudimentariamente), por ejemplo, la dinámica de las neuronas biológicas" (Rubio, Giri & Ilcic 2023, p. 148).

La comparación entre una computadora y el cerebro (humano específicamente) no es nueva y resulta curioso que esta comparación goza de bastante popularidad incluso entre personas que no tienen el menor indicio de cómo funciona el uno ni el otro. Podemos decir, sin embargo, que tal comparación es, cuando menos, intuitiva: ambos funcionan con impulsos eléctricos, tienen entradas de información y un procesador que reelabora esa información entrante (*inputs*) en respuestas salientes (*outputs*).

Pero no, nuestro cerebro, ni el de ningún otro animal conocido, funcionan como una computadora. Hay algunos ingredientes básicos fundamentales y absolutamente necesarios para entender nuestro cerebro: una compleja historia evolutiva de millones de años, los sentimientos, las emociones, el aprendizaje constante producto de la relación de su portador con el entorno, las predicciones no programadas y muy significativamente los procesos de conciencia (Mora 2020).

Los ya citados Hawkins y Blakeslee enumeran también algunas de las diferencias fundamentales:

> Las computadoras y los cerebros se basan en principios completamente diferentes. Uno está programado, el otro es autodidacta. Uno tiene que ser perfecto para trabajar, uno es naturalmente flexible y tolerante a los fracasos. Uno tiene un procesador central, otro no tiene un control centralizado. La lista de diferencias es interminable (Hawkins & Blakeslee 2004 p. 12).

Como bien sugiere Mora, si bien la computadora procesa información (como también lo hacen los cerebros), tal información es procesada sin "saber" que se está procesando. Por su parte, el cerebro humano, con la conciencia, "sabe" lo que hace (al menos a veces).

> el ser humano poseedor del cerebro que procesa toda información no ve ni oye ni percibe nada (a pesar de estar rodeado y bombardeado constantemente por todos los estímulos sensoriales que le rodean) a menos que aquella información sensorial tenga algún significado para él. Solo ante aquello que significa algo la maquinaria atencional del cerebro se pone en marcha. Sólo cuando se tiene hambre, el alimento significa algo y se detecta en el entorno rápidamente. Precisamente para detectar el alimento y previo a ello hay que tener hambre, es decir, hay que poner en marcha la maquinaria emocional que es la que detecta informaciones sensoriales que dicen algo. Es entonces cuando el cerebro se pone a trabajar y procesa la información sensorial correspondiente (Mora 2020, p. 52).

Pinker se suma a los críticos de la analogía. Si bien afirma que "pensar es computar" sostiene que la "metáfora del ordenador" es problemática porque mientras las computadoras son rápidas, los cerebros son lentos: "Las partes que componen un ordenador son fiables, las que componen el cerebro llevan ruido incorporado. Los ordenadores cuentan con un número limitado de conexiones, los cerebros tienen billones. Los ordenadores son montados siguiendo unos planos, los cerebros deben montarse solos" (Pinker 2008, p. 46). Churchland (1990), por su parte, dirá que la comparación es confusa, equívoca y

perniciosa y que comparar al cerebro y la mente con un ordenador y su software respectivamente es un dualismo innecesario y un error.

4. Conclusiones

Si bien la inteligencia permite a los humanos y otros animales resolver problemas, es más que eso. Hay varios componentes básicos que permiten resolver esos problemas de manera eficiente y flexible, siendo dos de los más importantes la memoria y la predicción. No puede haber predicción sin memoria.

Desde los debates generados en torno a la inteligencia artificial, en no pocas ocasiones se obvia un atributo esencial de la inteligencia humana (y de otros animales) a la hora de extrapolar el término inteligencia: nuestra inteligencia es flexible y situacional. Aunque fallemos constantemente en nuestras tareas de resolución de problemas y aunque no seamos hiperexpertos en algo (como sí lo son los sistemas de IA), nuestro cerebro se acomoda al contexto y a la situación. Busca salidas y, de manera no consciente, hace predicciones que nos mantienen con vida.

La inteligencia, como la conocemos, se mide por la capacidad de recordar y predecir patrones del mundo y está alimentada por lo que ese mundo le proporciona, incluido el mundo social, las emociones y la cultura. Si queremos seguir sosteniendo que la inteligencia artificial es inteligente, habría entonces que reconsiderar el concepto mismo de inteligencia.

Repensar la inteligencia desde una perspectiva más amplia implica reconocer que no se trata únicamente de procesamiento de información o rendimiento en tareas específicas, sino de una integración dinámica entre el organismo y su entorno. La inteligencia emerge de una interacción continua entre memoria, predicción, experiencia corporal, entorno físico y social. Esta comprensión más holística resalta las limitaciones de equiparar la inteligencia artificial con la humana, al carecer aquella de cuerpo, emociones, historia evolutiva y una inserción genuina en contextos vitales. Solo reconociendo estas diferencias podremos avanzar hacia definiciones más rigurosas y útiles del término "inteligencia".

Bibliografía

Bardi, L., Gheza, D. y M. Brass (2017), "TPJ-M1 Interaction in the Control of Shared Representations: New Insights from tDCS and TMS Combined", *Neuroimage* 146: 734-740.

Byrne, R. (1995), *The Thinking Ape: Evolutionary Origins of Intelligence*, Oxford: Oxford University Press.

Chollet, F. (2017), "The Implausibility of Intelligence Explosion", *Medium*. https://medium.com/@francois.chollet/the-impossibility-of-intelligence-explosion-5be4a9eda6ec

Churchland, P. (1990), *Neurophylosophy. Towards a Unified Science of the Mind-brain*, Cambridge: MIT Press.

Colmenares, F. (2015), *Fundamentos de psicobiología. Volumen 2: Comportamiento y procesos psicológicos en contexto evolutivo*, Madrid: Editorial Síntesis.

Crawford, K. (2024), *Atlas de inteligencia artificial. Poder, política y costos planetarios*, Bogotá: Fondo de Cultura Económica.

Dreyfus, H. (1972), *What Computers Can't Do: A Critique of Artificial Reason*, New York-Evanston-San Francisco-London: Harper & Row, Publishers.

Frith, C. y U. Frith (2023), *What Makes Us Social?*, Cambridge: MIT Press.

Gopnik, A. y A. Meltzoff (1999), *Palabras, pensamientos y teorías*, Madrid: Visor.

Gopnik, A., Meltzoff, A. y P. Kuhl (1999), *How Babies Think. The Science of Childhood*, London: Weidenfeld & Nicolson.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G. y R. Blake (2000), "Brain Areas Involved

in Perception of Biological Motion", *Journal of Cognitive Neuroscience* 12: 711-720.

Hawkins, J. y S. Blakeslee (2004), *On Intelligence. How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*, Nueva York: Times Books, Henry Holt and Co.

Hawkins, J. (2023), *Mil cerebros. Una nueva teoría de la inteligencia*, Barcelona: TusQuets Editores.

Julia, L. (2019), *L'intelligence artificielle n'existe pas*, Paris: Éditions First.

Larson, E. (2022), *El mito de la inteligencia artificial. Por qué las máquinas no pueden pensar como nosotros lo hacemos*, Barcelona: Shackleton Books.

Llinás, R. (2001), *I of the Vortex: From Neurons to Self*, Cambridge: The MIT Press.

Madrid Casado, C. (2024), *Filosofía de la inteligencia artificial*, Oviedo: Pentalfa Ediciones.

Miele, D., Wager, T., Mitchell, J. y J. Metcalfe (2011), "Dissociating Neural Correlates of Action Monitoring and Metacognition of Agency", *Journal of Cognitive Neuroscience* 23: 3620-3636.

Mora, F. (2020), *Cómo funciona el cerebro*, Madrid: Alianza Editorial.

Pelphrey, K., Singerman, J., Allison, T. y G. McCarthy (2003), "Brain Activation Evoked by Perception of Gaze Shifts: The Influence of Context", *Neuropsychologia* 41: 156-170.

Perazzo, R. (1998), *De cerebros, mentes y máquinas*, Buenos Aires: Fondo de Cultura Económica.

Pinker, S. (2008), *Cómo funciona la mente*, Barcelona: Editorial Destino.

Rubio, E., Giri, L. y A. Ilcic (2023), "Desafíos epistemológicos en la era de las redes neuronales artificiales: abordando sistemas complejos desde una perspectiva computacional", *Argumentos de Razón Técnica* 26: 145-178.

Rumelhart, D., Hinton, G. y R. Williams (1986), "Learning Representations by Back-propagating Errors", *Nature* 323: 533-536.

Sigman, M. y S. Bilinkis (2024), *Artificial. La nueva inteligencia y el contorno de lo humano*, Bogotá: Debate.