

# LA INTELIGENCIA ARTIFICIAL Y LA CONQUISTA DE LO MENTAL

**Mariela Destéfano**

---

UBA

*El tema: capturar a la mente*

¿Logró “Deep Blue” ganarle un partido de ajedrez al campeón mundial Kasparov debido a que pensaba? Esta pregunta se resiste a tener una única respuesta. Quienes se aferran a las intuiciones más chauvinistas acerca de lo mental se niegan a dar una respuesta positiva y consideran que son las personas los entes prototípicamente pensantes. Sin embargo, la respuesta más interesante será la de quienes confían en que efectivamente “Deep Blue” puede pensar. Si se entiende lo mental como un cómputo de representaciones simbólicas y a su vez sabemos que máquinas como “Deep Blue” realizan este procesamiento computacional dirigido por los programas de instrucciones cargados en sus memorias, entonces es fácil concluir que estas máquinas literalmente tienen mente. La anterior afirmación es la que se toman en serio programas clásicos de investigación como la inteligencia artificial ( I.A.) fuerte, disciplina que, junto con la psicología cognitiva, la lógica y la lingüística, entre otras, conforman una ciencia cognitiva unitaria cuyo objetivo general es abstraer las características comunes de una nueva clase de cosas cognoscentes que incluye no sólo a los vertebrados superiores, sino también a algunos sistemas informáticos. Las respuestas a la pregunta planteada inicialmente dejan translucir cuál es la polémica de fondo. En conexión con el tradicional problema de la relación mente- cuerpo, tomar posición respecto de la cuestión de que una máquina piense significa haber respondido a la pregunta metafísica por la mente. Estimo que no es posible que un ordenador como “Deep Blue” piense, mas no por considerar que lo mental se circunscriba al ámbito de los seres biológicos, y más exclusivamente humanos, sino, porque admitiendo la alternativa propuesta por la I.A. fuerte se estaría aceptando una idea sesgada acerca de lo que es la mente. Mi hipótesis es que las máquinas concebidas bajo el paradigma clásico de la I.A. no pueden pensar dado que éste no atiende a una noción acabada de la mente, donde la misma se explique no sólo por el procesamiento de información representacional desencadenado ante un estímulo y productor de una respuesta, sino que también por su relación con factores externos tan

activos como dichos procesos internos. Sólo una I.A. que advierta esta particularidad de lo mental podrá idear máquinas pensantes.

En *el juego de los que piensan* desarrollaré brevemente el juego de la imitación gracias al cual Turing concluye que las máquinas piensan y expondré cómo es que a partir del mismo se desprenden los elementos conceptuales básicos acogidos por la I.A. fuerte. En *hacia una nueva noción de lo mental* explicitaré los problemas que acarrea un juego conductual como el de Turing para concluir que hay cierto tipo de conductas, pertenecientes a la esfera relacional de la mente, que no son correctamente evaluadas en dicho juego. Es esta falla, que evidencia la escasa atención que se ha prestado a la mente relacional a la hora de diseñar el juego de la imitación, la que impide inferir que las máquinas tienen mente. Acabada la *pars destruens* del trabajo edificaré la *pars construens* defendiendo el aspecto relacional del pensamiento a través de un argumento que atienda al carácter adaptativo de la mente. Finalizaré con la exposición de una serie de cuestiones que debería contemplar una I.A. mejorada que se proponga imitar una mente que presenta un carácter adaptativo. Será en la conclusión donde retome los resultados obtenidos.

### *El juego de los que piensan*

En "Computing Machinery and Intelligence" (1950) Turing lleva a cabo la defensa de la tesis de que las máquinas poseen mente mediante su famoso "juego de la imitación". El mismo consta de tres participantes: un interrogador frente a un hombre y una mujer escondidos en habitaciones separadas. El primero de ellos desconoce el sexo de los demás jugadores y su objetivo es determinar cuál es el hombre y cuál la mujer a través de la formulación de ciertas preguntas. Para que el interrogador no descubra el sexo de sus interrogados debido al tono de sus voces, las respuestas al cuestionario pueden hacerse por escrito, mediante un teleimpresor o con la ayuda de un intermediario. Lo cierto es que si una máquina toma el lugar del participante masculino, por ejemplo, pudiendo la misma imitar las respuestas que daría este jugador escondido, entonces no se vería alterada la decisión que tome el interrogador respecto de cuál de los participantes es el hombre. Y si la máquina pasa el test, gracias a una impecable imitación de las manifestaciones cognitivas humanas, entonces no hay razón alguna para negar que piensa. Pero ¿qué nos autoriza a admitir que una máquina reproduzca perfectamente las respuestas humanas? Aquí se evidencia que, si bien es cierto que al

estudiar la posibilidad de que las máquinas piensen Turing pretende evitar especulaciones subjetivas acerca de lo que significa pensar apelando a un método objetivo como lo es el juego presentado, también es verdad que esta estrategia sólo es posible bajo la aceptación previa de un supuesto básico que clarifique qué se ha de entender por mente. Este supuesto identifica al pensamiento con el resultado explícito obtenido como consecuencia del procesamiento de información representacional puesto en marcha ante un determinado estímulo. Bajo este supuesto la máquina de Turing puede reproducir la actividad pensante ya que el hecho de que ella también procese datos mediante un manejo computacional de las representaciones simbólicas, la habilita para generar efectos conductuales semejantes a los del hombre. Dicha capacidad de cómputo no encuentra límite alguno y es ésta la universalidad propia de una máquina que no se la concibe como un ordenador material concreto, sino que como una ideación de carácter abstracto.

La I.A. fuerte es el programa de investigación que rescata la concepción acerca de lo mental manejada por Turing. El procesamiento de información de los ordenadores modernos, construidos en el seno de dicho proyecto investigativo, está regido por un programa definido como una secuencia de instrucciones que indican qué acciones deben llevarse a cabo según las condiciones que le son antecedentes. De este modo, si la actividad computacional que desemboca en resultados explícitos es lo característicamente mental y si dicha actividad sólo es posible gracias a la regulación de un software determinado, entonces la posesión de tal programa es el que nos permite afirmar que la máquina tiene mente. La misma posee la universalidad de la máquina abstracta de Turing sólo en el sentido de que su memoria siempre puede ampliarse como para incorporar nuevos programas que pauten nuevos cálculos anteriormente irrealizables.

### *Hacia una nueva noción de lo mental*

#### a- Máquinas que no piensan

Por lo visto, en el juego de la imitación lo que adquiere relevancia para el interrogador son las respuestas que dan sus interrogados. Así, se puede afirmar que el test que el primero lleva a cabo para la determinación del sexo de los jugadores tiene un carácter indiscutiblemente conductista, pues por el mismo se considerará que aquel participante oculto que, por ejemplo, se comporte como una mujer, será efectivamente

una mujer. Teniendo esto en claro puede decirse que la maniobra de reemplazar al hombre por la máquina es una forma de responder a la pregunta : “¿ Saben pensar las máquinas?” de manera conductista. De este modo, una máquina que se comporta como teniendo mente - ya que a partir de los inputs ambientales y un proceso consecuente da respuestas conductuales indistinguibles de las de un ser pensante como el hombre-, efectivamente piensa. Pero veamos qué problemas genera la evaluación de lo mental a partir de los parámetros conductistas adoptados en el juego de la imitación.

i) Es cierto que la máquina debe poder imitar todo tipo de conducta. Mediante sus preguntas el interrogador aspira a evaluar cualquier comportamiento humano, sin embargo con su test también podría evaluar conductas no humanas y en este caso la máquina debería imitar tales comportamientos. Imaginemos que en vez de un hombre y una mujer hay dos perros ocultos. El interrogador es el dueño de uno de ellos, al cual ha entrenado de tal manera que, ante determinados sonidos de su voz, responde con determinada cantidad de ladridos. El otro perro también responde a la mayoría de estas instrucciones, pero no a toda de ellas. El objetivo del interrogador es descubrir cuál de los dos perros es el suyo. Mas, si reemplazamos a su perro por una máquina programada para emitir los mismos ladridos, el interrogador podrá concluir que determinado jugador es su perro cuando en realidad era una máquina. Lo cierto es que no estaríamos dispuestos a sostener que una exitosa reproducción de las conductas caninas vuelvan a nuestra máquina en una máquina pensante. El error de un juego que presenta un carácter eminentemente conductista es considerar que la imitación es la base de la atribución de mentalidad, pues a partir de ello se llegaría al absurdo de que cualquier cosa que pueda imitar tendrá mente. Así, el juego probaría demasiado y se volvería altamente sospechoso.

ii) Pero aún aceptando que la imitación es la base para la atribución de mentalidad, el juego presenta un problema insalvable. En la primera sección del trabajo hemos clarificado que el éxito que tienen las máquinas en la imitación de la conducta humana está dado por el hecho de que tanto en ellas como en el hombre las respuestas conductuales resultan de un previo procesamiento interno de información representacional iniciado por un determinado estímulo. Si en ello radica el pensamiento y las máquinas pueden emularlo, entonces las mismas efectivamente piensan. La dificultad se origina cuando la máquina intente reproducir con fidelidad acciones que requieran de un procesamiento donde se vuelva necesario algún tipo de contacto con el mundo. Supongamos que tanto la mujer como la máquina están encerrados en habitaciones

separadas y cada una de ellas cuenta con una máquina expendedora de gaseosas. Sólo el interrogador sabe que depositando una moneda de un peso y apretando el botón rojo, la gaseosa expendida es una Sprite. Una de sus preguntas, sabiendo que sus interrogados cuentan con la moneda de un peso, es la siguiente: “¿ qué gaseosa se obtiene introduciendo una moneda de un peso y apretando el botón rojo de la máquina expendedora de gaseosas?” La mujer, luego de realizar la maniobra responde “Sprite”. Teniendo en cuenta que la máquina en cuestión en su carácter abstracto tiene una capacidad de cómputo universal e irrestricta, la misma evaluando las distintas respuestas programadas ante un plexo de determinadas situaciones preconcebidas, responderá “Sprite”. Podríamos sostener con Turing que, dado que la máquina se comportó tal como el humano, entonces tiene una mente. Pero aunque máquina y humano dieron la misma respuesta cada uno puso en juego distintos procedimientos para su elaboración: la primera encaró un procedimiento computacional de naturaleza intrínseca, el segundo, un procedimiento que incluye además relaciones con elementos externos. Parecería que este test conductual es necesario pero insuficiente a la hora de determinar si algo tiene mente. Su efectividad estaría garantizada si lograrse también evaluar en su especificidad este tipo particular de conducta humana donde el procesamiento de los datos se hace en relación con factores externos. Pero ello haría que el juego de la imitación presuponga una noción nueva y más rica de lo que es la mente, ya que tales conductas pertenecen a un tipo de *mente relacional* diferente de la *mente computacional* tenida en cuenta por Turing.<sup>1</sup> Así, de acuerdo a lo dicho puede extraerse el siguiente *Tollens*:

- Premisa mayor: *Si el juego de la imitación verdaderamente muestra que una máquina piensa, entonces la mente es sólo la conducta explícita desencadenada por el procesamiento interno de representaciones simbólicas iniciado por un estímulo ambiental.*

- Premisa menor: *Pero no es cierto que la mente sólo sea la conducta explícita desencadenada por el procesamiento interno de representaciones simbólicas iniciado por un estímulo ambiental*

---

<sup>1</sup> En esta distinción de tipos de mentalidad no quisiera dar a entender que los estados de las máquinas carezcan de toda nota relacional, pues lo cierto es que en su carácter funcional son estados que se relacionan con todo el resto de los estados de la máquina. Pero esta es una relación interna, y la razón por la cual denomino “mente relacional” al modelo de mente que defenderé es que el carácter relacional se hace más explícito al intervenir elementos externos al sistema. Es cierto que la anterior es una mera diferencia de grado, dado que es inevitablemente cierto que todo sistema computacional está en relación directa con su medio. Con todo, en este trabajo sólo intento mostrar lo que queda por fuera de la propuesta de Turing, sin detenerme en establecer asociaciones- que efectivamente existen- entre los dos modelos de mentalidad analizados. Quise atender a aquellos estados para los cuales el medio ambiente no juega un mero papel de input incitador de procesamientos internos, sino que tiene el mismo status que los cálculos internos.

- Conclusión: *No es cierto que el juego de la imitación verdaderamente muestre que una máquina piensa.*

Con ello quedaría desestimada una I.A. que adscriba a la propuesta teórica de Turing.

### *Mente y adaptación*

El argumento anterior exige una justificación de la premisa menor, pues el ejemplo de la máquina expendedora de gaseosas constituye una aproximación intuitiva en la defensa de tal premisa. Empezaré dicha justificación a través de otro argumento donde se ponga en relieve el carácter adaptativo de la mente:

- *La mente constituye un rasgo adaptativo del hombre.*

Este es un dato que nos brinda la teoría evolucionista. Es cierto que naturalistas como Wallace consideraban que la selección natural sólo puede dar cuenta de habilidades prácticas que favorezcan la supervivencia dejando sin explicar capacidades superiores como el pensamiento. Pero aquí sostendremos con Darwin que dichas capacidades de alguna manera también son producto de la selección natural, convirtiéndolas en rasgos adaptativos para el hombre.

- *La noción de adaptación es una noción eminentemente relacional.*

Es incorrecto entender la idea de adaptación a través de la imagen de un “ambiente que presenta problemas y un organismo que da las soluciones”. Plantear la *adaptación* del organismo al medio es plantear una *relación* de interacción simétrica entre ambos componentes, donde cada uno de ellos está abierto al otro modificándose en una mutua influencia. Es debido a esta comunión que no puede concebirse un organismo sin ambiente ni un ambiente sin organismo. Pues, las características que hacen al organismo las ha adquirido en virtud de su relación con el ambiente, así como también las características que definen a un ambiente sólo son aquellas condiciones físicas relevantes para la supervivencia del organismo.

- *La mente constituye un rasgo relacional del hombre.*

Este enunciado resulta de substituir en la primera premisa la noción de adaptación por la de relación, reemplazo posible gracias a la introducción de la segunda premisa.

Siguiendo la idea de adaptación explicada, el carácter relacional de la mente se hace patente en ciertas acciones que son el resultado de un tipo de procesamiento previo que apela a la interacción simétrica entre las aptitudes propias del individuo<sup>2</sup> y factores externos. Esta simetría hace que dichos factores tengan un rol tan activo como el de las aptitudes del sujeto a la hora de operar modificaciones. Dado que el proceso en el cual tiene lugar esta dialéctica transformacional posee un valor cognitivo para el individuo, ya que es un modo mediante el cual el mismo aprende, entonces, las *acciones* que produzca también tendrán un *carácter cognitivo*<sup>3</sup> Debido a la compenetración existente entre aptitudes individuales y factores externos no puede concebírseles con independencia. Las primeras están determinadas por su relación con el medio, así como también el medio debe ser entendido exclusivamente como aquellas propiedades manipuladas por las aptitudes del sujeto para llevar a cabo el proceso cognitivo deseado.

*- A partir de lo anterior se desprende que no es cierto que la mente sólo sea la mera conducta desencadenada por el procesamiento interno de representaciones simbólicas iniciado por un estímulo ambiental.*

Lo cierto es que dado que la mente evidencia la naturaleza relacional que hemos presentado, entonces no puede ser explicada por una mera caracterización computacional donde no se tenga en cuenta la participación de elementos extrínsecos al sistema. Si tomamos el juego de la imitación tal como lo expone Turing y a su vez aceptamos esta nueva idea acerca de lo mental, sólo se demostrará que quien piensa es el interrogador. Esto es así ya que el método de preguntas que él lleva a cabo es parte de un proceso cognitivo donde sus aptitudes entran en interacción con factores externos activos, como lo son las respuestas de sus interrogados. El resultado de tal proceso es la acción cognitiva mediante la cual asigna los sexos a los jugadores. Todo ello evidencia que quien estuvo en una actitud relacional fue el interrogador y no la máquina, por lo que sólo al primero podrá adscribirsele mentalidad.

Para que podamos decir que las máquinas piensan debemos idearlas satisfaciendo el aspecto adaptativo-relacional propio de la mente. Pero la imitación de

<sup>2</sup> Posiciones como las de Clark las entienden como características físicas de un agente, y es por eso que defienden la idea de “an embodied mind”. En este trabajo hablo de “aptitudes” para evitar con esta terminología tan general tomar posición respecto de la naturaleza de aquello que interactúa con el medio. Afirmar que estas aptitudes son físicas/ biológicas me trasladaría al problema mente- cuerpo que no pretendo tematizar.

<sup>3</sup> Acciones cognitivas como las movidas realizadas en el Scrabble y el Tetris son ejemplos estudiados por Kirsh y Maglio- véase Clark, A., Chalmers, D.; 1998- en el área de las ciencias cognitivas, que son frecuentemente citados en la bibliografía. No es mi objetivo hacer una caracterización exhaustiva de dichas acciones. Para los fines de mi trabajo me basta con mostrar su existencia a partir del concepto de adaptación.

esta mente con carácter adaptativo plantea ciertas cuestiones que debe tener en cuenta una I.A. mejorada que emprenda este nuevo camino investigativo. En primer lugar, puede entenderse sincrónicamente que la mente es un rasgo adaptativo que los humanos poseen en este momento evolutivo, el cual resulta favorable ya que aumenta la eficacia de sus portadores. Sin embargo, aunque bajo esta comprensión del carácter adaptativo de lo mental se evite el supuesto inverificable de que en el pasado hubo selección natural del rasgo, la I.A. debe clarificar cómo es que la posesión de una mente aumenta la eficacia de una máquina, y, por sobre todo, qué debe entenderse por eficacia artificial, cuando sabemos que la de eficacia es una noción biológica propia de la teoría evolucionista. Si, por el contrario, consideramos a lo mental diacrónicamente como una adaptación producto no de un proceso azaroso sino de la acción pasada del mecanismo de selección natural, entonces, aunque la I.A. no tenga que reproducir la historia evolutiva del rasgo mental, sí deberá construir máquinas pensantes lo suficientemente flexibles como para que se incluyan en una dinámica evolutiva. Aquí debe aclararse en qué sentido puede hablarse de evolución artificial. Por otra parte, si interpretamos que la mente como rasgo adaptativo no es producto directo de la selección natural, sino que es un subproducto de otro rasgo del organismo seleccionado naturalmente, entonces la I.A. debería identificar a este último y reproducirlo en las máquinas para que en ellas se de el rasgo mental consecuente.

Por lo visto una I.A. que preste atención a la naturaleza relacional de la mente en su objetivo de crear ordenadores pensantes, deberá estar al tanto de los avances producidos en el área de la biología. Este camino de investigación parece tener buen futuro en áreas de la inteligencia artificial donde se profundiza en el estudio de algoritmos genéticos y de sistemas inteligentes autónomos con aprendizaje inductivo por interacción con el medio. Sin embargo excede los límites de mi trabajo profundizar en dichas cuestiones.

### *Conclusión*

A lo largo de este trabajo he intentado mostrar que la noción de mente que la I.A. fuerte toma de Turing es limitada. Hemos visto que el juego de la imitación por el cual dicho autor pretende concluir que las máquinas piensan constituye un criterio conductual de evaluación del pensamiento. Esto acarrea dos problemas. En primer lugar, aceptando que la máquina reproduce todo tipo de conducta, puede admitirse que su capacidad de

imitación también incluye los comportamientos no humanos, en cuyo caso no estaríamos autorizados a concluir que la máquina piensa y se evidenciaría el error de considerar que la imitación es condición suficiente para la atribución de mentalidad. En segundo lugar, aún concediendo que la imitación es condición suficiente para la atribución de mentalidad, habrían ciertas conductas humanas que la máquina no podría emular con fidelidad. Esto es así porque dichas conductas son explicables bajo una concepción distinta de lo mental de la que es tenida en cuenta en el juego de la imitación. Para abordarlas ha de modificarse la idea de que la mentalidad consiste en la generación de respuestas conductuales causadas por un procesamiento interno de representaciones simbólicas, desencadenado por un estímulo ambiental. Es necesario incorporar activamente a los elementos del mundo como elementos reveladores a la hora de responder la pregunta por la mente. Esta concepción de mentalidad queda justificada por el hecho de que la misma es un rasgo adaptativo humano y como la adaptación es una noción relacional donde interactúa un organismo con su ambiente, de allí se sigue que la mente también tenga este carácter relacional. Lo cierto es que la I.A. sólo logrará crear máquinas pensantes atendiendo a esta noción de lo mental. Teniendo en cuenta que aquí se subraya que la mente es un rasgo adaptativo humano la I.A. deberá resolver ciertas cuestiones como qué significa hablar de eficacia y evolución artificiales, sin perder de vista los aportes de la biología evolucionista.-

#### Bibliografía

- Churchland, P. M. (1999), *Materia y conciencia*, Barcelona, Gedisa.
- Clark, A., Chalmers, D., (1998), "The Extended Mind", *Analysis* 58,10-23.
- Clark, A.; (1997) *Being There*, Massachusetts, The MIT Press.
- García Martínez - Servente, Pasquini (2003), *Sistemas Inteligentes*; Buenos Aires, Nueva Librería.
- Lewontin, R. C., *Genes, Organismo y ambiente*, Barcelona, Gedisa.
- Lewontin, R. C., (1984), *No está en los genes*.
- Rivière, A, (1991), *Objetos con mente*, Madrid, Alianza.
- Sober, E. (1996), *Filosofía de la Biología*, Madrid, Alianza.

Toribio, J. (2000) "Internalismo, externalismo y ecología" en Chacón Fuentes, P., Rodríguez Gonzales. M.; *Pensando la mente*; Madrid, Biblioteca Nueva.

Turing, A. M. (1950) "Los aparatos de computación y la inteligencia" en Hofstadter, D., Dennett, D. C., *El ojo de la mente*, Buenos Aires, Sudamericana.