

¿ES CONSTRUIBLE LA INTELIGENCIA?: LA ELABORACION DE TEORIAS EN LA SIGUE UN CAMINO INVERSO RESPECTO A LAS CIENCIAS EMPIRICAS

M.L.MARQUINA

Univ. Politécnica, Facultad de Informática, Madrid.

En las últimas décadas se ha polemizado mucho, quizá en exceso, acerca de si las máquinas, como manipuladoras de símbolos, pueden ser consideradas como un modelo restringido de la mente humana. Una interpretación tal de la mente, en términos de máquina pensante, es restringida porque no pretende instanciar todas las posibles variables que desencadenan la conducta inteligente ni, tampoco, todos los posibles tipos de inteligencia. La relación mente-máquina ha hecho correr ríos de tinta en ámbitos diversos pero interrelacionados, tales como la psicología cognitiva, la filosofía de la mente o la inteligencia artificial (IA). De hecho, los orígenes de esta última disciplina en particular fueron los que aportaron la chispa necesaria para que la teorización sobre la mente comenzase a correr por nuevos derroteros, tanto en filosofía como en psicología, de forma especial. Esta situación ha hecho que un tema tradicional recurrente, como es el de la relación mente-cuerpo, haya sido en gran parte sustituido por el del posible vínculo entre mentes y máquinas. Esto, en particular, permitirá factorizar la inteligencia, ya que se entiende que tener mente es una condición necesaria -desconocemos si suficiente- para mostrar inteligencia. La que posean las máquinas y, sobre todo, la clase a la que pertenezca, dependerá de si tienen o no mente y, si la tienen, en qué sentido.

En una breve mirada retrospectiva, la mente puede ser vista como una hipótesis de trabajo con la que algunas teorías antropológicas o psicológicas han intentado justificar las ventajas adaptativas de los seres humanos. Según esto, poseer mente supone ganar terreno en el proceso de selección natural porque la inteligencia en ella constreñida, fruto del desarrollo evolutivo, permite a los que la poseen superar mejor la supeditación a la naturaleza. Hay quienes piensan que es la misma naturaleza quien

"...ha decidido...que evolucionen seres sentientes como nosotros mismos en lugar de contentarse con criaturas que podrían estar

Endoxa: Series Filosóficas, nº 1, 1993, UNED, Madrid:

M.L.MARQUINA, ¿Es construible la inteligencia?: La elaboración de teorías en IA sigue un camino inverso respecto a las ciencias empíricas;

pp. 393-400

dirigidas por mecanismos de control completamente inconscientes. Si la conciencia no sirve para ningún propósito evolutivo, ¿por qué la Naturaleza se tomó la molestia de hacer evolucionar cerebros conscientes cuando parece que hubieran bastado cerebros 'automatas' no-sentientes como los cerebelos?" [Pen91, p.181].

Suena grandilocuente esto. Mente y conciencia han de ir unidos para que aflore la 'inteligencia real'; además, ello constituye una importante decisión de la naturaleza. Sin embargo, las palabras de Penrose son consistentes con el 'principio antrópico' al que él mismo se refiere. Según éste principio es la naturaleza del universo en el que estamos la que está determinada por la exigencia fundamental -no se sabe para quién- respecto a que tengan que existir seres que sienten, como nosotros, para observarlo. Lo que parece claro es que este punto de vista constituye una hipótesis que conjuga la investigación con la creencia acerca de cómo parece que deban ser las cosas. Pero las cosas, simplemente, son, ocurren, y el cómo es lo que la ciencia tiende a explicar. Es difícil alcanzar una construcción neutral y objetiva del conocimiento, pero la ciencia debe perseguirla guiada por el tipo de explicaciones en el que se asienta, y que es lo que básicamente la distingue de otras formas de conocer. A fin de cuentas, todas las formas de teorizar suponen crear modelos que se forjan sobre relaciones entre estructuras conceptuales y dominios de interpretación; es lo que se pretende explicar con esas relaciones lo que distingue el discurso científico del religioso, por ejemplo. Uno de los principales problemas que plantea la teorización sobre la mente es que se la analiza desde vertientes demasiado diversas, tan diversas que, en ocasiones, se vierten explicaciones que pueden ser elaboradas inconscientemente desde disciplinas cuyas ontologías son incompatibles. Ni que decir tiene que ello da lugar a un importante confusiónismo intelectual. A lo mejor las palabras de Penrose pueden sugerir algo al respecto. Al margen de la argumentación precedente, es importante señalar que para la teoría de la mente la IA es una fuente significativa de nuevos argumentos que pueden contribuir a precisar, si no el sentido de la mente si, al menos, la utilidad que supone poseerla. De hecho, los desarrollos en IA dedican sus principales esfuerzos a instanciar los tipos de inteligencia 'útil' (como si hubiese alguno realmente inútil) con los que realizar tareas en forma eficiente. Ello supondrá elegir un lenguaje cuyos términos nombran y representan determinados aspectos del procesamiento humano de la información, y un dominio que estará constituido por rasgos caracterizadores de un determinado tipo de inteligencia. La mente puede pertenecer a la ontología de ese dominio, siempre que sea considerada como una condición formal necesaria para la realización de tareas con dosis de inteligencia, si es que se puede ser inteligente a medias. Precisamente la formalidad es un

requisito básico de las teorías basadas en la hipótesis del lenguaje del pensamiento defendidas por autores como Fodor y Phylysin cuando argumentan en disfavor de las aproximaciones conexionistas a la mente. Según dicha hipótesis, el pensamiento requiere un sistema representacional que es similar al lenguaje en sus aspectos más relevantes, y ese sistema representacional es dependiente de la condición de formalidad que, hasta el momento, sólo ha sido instanciada en las computadoras digitales clásicas [Fod88].

Lo cierto es que mientras las líneas de investigación sobre la mente se han incrementado en número en los últimos tiempos en direcciones bien distintas (pues la relación mente-máquina se puede establecer con diferentes clases de máquina, fundamentalmente, programables y no programables; es decir, máquinas tipo Turing y redes neurales), las que se han llevado a cabo respecto a la inteligencia se han mantenido en una dirección mucho más uniforme, y ello a pesar de que cuando hablamos hoy de la inteligencia, aún no se sabe bien de qué se está hablando realmente. Se sabe que existe una variedad de pruebas que tienen el objetivo fundamental de medir lo que ella sea; y, sobre todo, se cuenta con una visión intuitiva, si no de lo que ella es, si de lo que supone poseerla. Esto es lo que permite a menudo a las personas, aplicando el nada simple razonamiento de sentido común, percatarse de la presencia de la estupidez o de la inteligencia, aun no siendo en absoluto capaces de definir o incluso caracterizar lo que esta última es. Simons, por ejemplo, apunta algunas ideas generales respecto a la representación que generamos de los procesos inteligentes:

"Solemos imaginar que la inteligencia consiste en la habilidad para resolver problemas, hacer sumas, aprender, salir airoso ante lo desconocido, etc. Solemos asumir que las personas en posiciones de responsabilidad -directivos, profesionales, ministros, etc.- son inteligentes. Tal suposición en ocasiones nos lleva a la siguiente paradoja: ¿si es tan inteligente, por qué él/ella actúa/habla tan estúpidamente?. Esto nos motiva a identificar las diferentes clases de inteligencia. Un brillante matemático puede no tener perspicacia política, y un especialista en Joyce puede carecer de tacto para las relaciones humanas" [SIM87, p.17].

Se trata de un texto jugoso, dado el conjunto de presuposiciones que en él se ven expresadas. No obstante, no vamos a polemizar sobre ellas porque lo que nos interesa en este momento es resaltar una de las hipótesis centrales de la IA: la racionalización de la inteligencia. Un poco jocosamente expresado, si se nos permite, ello conduce a considerar que en un sistema dado pueden convivir la inteligencia y la estupidez a la perfección.

Pero debe haber algún elemento 'ad-hoc' en la inteligencia de las personas que la posean porque no parece que sea reducible a la realización de tareas específicas o al desarrollo y perfeccionamiento de un grupo específico de habilidades con las que mejor llevarlas a cabo. Precisamente, la IA tiende a restringir el dominio de operatividad de la inteligencia como capacidad, utilizándola como una mera estrategia de resolución de problemas. Por ejemplo, T. Winograd y F. Flores consideran que se puede explicar esa estrategia en términos de una toma racional de decisiones que deberá involucrar, al menos, la caracterización del contexto de la tarea, el diseño de la representación formal, la incorporación de la representación en la computadora y la realización de un proceso de búsqueda. Todo esto supone analizar la realización de una tarea

"...en términos de un 'espacio de problema', generado por un conjunto finito de propiedades y operaciones. Una 'solución' es un punto especial de ese espacio que posee las propiedades deseadas. La resolución de problemas es un proceso de búsqueda a partir de una secuencia de operaciones que conducirán al punto de solución" [Win86, p.95].

Esta concepción de los procesos cognitivos o, por lo menos, de algunos de ellos (en términos generales, serían aquellos que involucran procesamiento simbólico más inteligencia de alguna clase), tiene que ver de forma directa con el matrimonio de conveniencia entre formalismo lógico y representación del conocimiento. Ello desencadena la puesta en acción de todo un conjunto de mecanismos automáticos de razonamiento en virtud del cual pueden inferirse conclusiones lógicas correctas a partir de hechos conocidos. Sin embargo, los sistemas que son eficaces en la realización de estas tareas son incapaces de comparar situaciones con base en razonamientos de tipo analógico. Es, en particular, en el despliegue de esas tareas donde no parece que las personas nos conduzcamos, representando primero en forma abstracta los problemas y, después, extrayendo conclusiones a través de razonamientos lógicamente válidos de las premisas en las que se condensa la representación abstracta. No hay que pensar -al menos con necesidad- que la naturaleza ha debido favorecernos dotándonos con la capacidad para interpretar los procesos cognitivos en forma tan abstracta. Además, no hay razón epistemológica (quizá sí técnica) para pensar que esa deba ser la mejor opción a la hora de interpretar la clase de procesos que subyacen a la conducta inteligente. Si se tiene esto en cuenta, con seguridad en el futuro se abrirán nuevas puertas que den paso a la elección de otras variables con las que construir nuevos modelos de la mente y, por tanto, también nuevas explicaciones de la inteligencia. Hasta el momento, la inteligencia ha sido construible si se toma en serio

la fragmentación de la misma, dando lugar a la construcción de sistemas altamente inteligentes y necios al mismo tiempo.

No obstante, en términos más generales que los de la IA parece que, en algún sentido importante, la inteligencia se puede concebir como una variable interviniente en la conducta de determinados sistemas biológicos. La presencia de esa variable daría lugar, en primer término, a una mayor independencia del medio y, en segundo y consecuentemente, a un conjunto no finito de posibilidades de acción no determinado de forma absoluta por las condiciones iniciales. Los sistemas que poseen esa variable, en cuanto capacidad, están en principio mejor dotados como especie para afrontar la competencia de la supervivencia. Este punto de vista hace mucho más difícil la tarea de construir teorías que funcionen como modelos adecuados de la inteligencia. En el proceso de elaboración de una teoría (para que ésta sea lo suficientemente buena con respecto al dominio al que se aplica, esto es, para que tenga el suficiente poder explicativo), se debería tener al menos una pista de cuáles son las variables más importantes que inciden en el fenómeno que se estudia. Esto genera incertidumbre en la investigación, y pone de manifiesto que la naturaleza gana la partida cuando no conseguimos explicar hechos conocidos ni predecir otros nuevos. Si cabe, podría considerarse que la naturaleza siempre 'sabe' más, incluso cuando competimos con ella con nuestras mejores teorías (por buenas que éstas sean, siempre puede quedar algo fuera). Evidentemente, han de entenderse estas palabras sin teñirlas de ningún tipo de vitalismo.

La cuestión importante sería poder conocer hasta qué punto podría ser alcanzable el objetivo de remedar el comportamiento inteligente en una forma lo más global posible; porque no se ha de olvidar que hasta el momento lo que se ha logrado es instanciar rasgos parciales de nuestra comprensión intuitiva de la inteligencia. Lo que falta es precisamente esa aproximación que de cuenta del mayor número posible de componentes de la misma. Pero la descomposición factorial de la inteligencia genera la necesidad de dar un salto que es en especial de índole cualitativa. Esto es, la capacidad de aislar rasgos de la inteligencia a través de la elaboración en IA de programas que los instancien, genera una comprensión fragmentada de los requerimientos de la inteligencia. Tocar fondo en esa comprensión global de la misma es ardua tarea, porque los mecanismos para la representación del conocimiento con los que cuenta la IA no parecen acercarse por el momento demasiado a la complejidad con la que la mente humana almacena los datos fragmentarios provenientes de la experiencia, fruto de su interacción con el entorno. Se podría señalar que no es una función lineal la que opera desde los fragmentos de inteligencia a la inteligencia global.

El estudio de la inteligencia en los últimos años, que vincula la psicología con la ciencia de las computadoras, ha tendido a hacer uso de

una estrategia metodológica denominada 'reducción del problema'. Dicha estrategia consiste, en la línea de lo dicho con anterioridad, en ramificar la consecución de un objetivo que, por su dificultad normalmente, no resulta accesible de manera directa. Así, los objetivos de largo alcance suelen ser perseguidos a través de un proceso de fragmentación en subobjetivos intermedios; la idea es que el cumplimiento de estos últimos conduzca a la consecución del objetivo principal que, en este caso, consiste en caracterizar de la forma más completa posible las restricciones que operan sobre la conducta de los sistemas inteligentes. Hofstadter señala que esta estrategia es un medio eficaz con el que convertir problemas globales en problemas locales. El problema fundamental consiste en reconstruir lo reducido, es decir, en pasar de los 'trozos de inteligencia' a simplemente la inteligencia. Luego lo que comenzó siendo la reducción del problema acabó por ser el 'problema de la reducción'. No se conocen bien las implicaciones que tienen las investigaciones en IA cuando lo que se trata de explicar es la operatividad de la inteligencia de un sistema como un todo, pues aplicando esa técnica de reducción a lo mejor lo que tenemos al final es un conjunto infinito de objetivos no particularmente cercano al objetivo principal del que se había partido. Es decir, quizá esta forma de afrontar el estudio de la inteligencia no alcance a explicar más que determinado tipo de procesos involucrados en ella y, además, pertenecientes a un sólo nivel de comprensión de la cognición: aquel para el que la IA puede construir un programa que los modele. Hofstadter, haciendo uso de un lenguaje adobado con ricas sugerencias manifiesta al respecto lo que sigue:

"...si uno fragmenta un problema dado en una serie de subproblemas, y luego divide éstos en subsubproblemas, y así sucesivamente, de manera recursiva, obtendrá al final objetivos sumamente modestos que, quizá, podrán ser satisfechos con un par de pasos. Al menos, así pareciera ser..." [Hof87, p.677].

Hofstadter quiere decir que la técnica de la reducción del problema puede llevar a generar un bucle infinito que haga del objetivo buscado una meta inalcanzable, tal y como les ocurrió a Zenón, a Aquiles y a la Tortuga. Puede no ser éste el caso cuando lo que se tiene entre manos no es cómo cubrir una distancia sino cómo caracterizar la inteligencia. De todos modos, al menos en forma tentativa, no estará de más tenerlo en cuenta por el momento.

Si sólo hubiésemos de tener en cuenta lo dicho hasta aquí, el proyecto de construcción de inteligencia habría quedado quebrado de inmediato. Y no es así porque se tiene la confianza de que la ciencia, que no es del todo una loca aventura, puede proporcionar modelos cada vez más refinados del

funcionamiento de la mente y, por tanto, de las propiedades de la inteligencia. En efecto, no se ha de olvidar que el siglo XX, al ver florecer el desarrollo de las computadoras, se ha encontrado con la posibilidad de concebir a las máquinas como modelos físicos de procesos abstractos. Ideas como ésta han ido forjando una visión 'desprendida' de la inteligencia y, en última instancia, una interpretación funcional de la mente que hace de la descripción del soporte físico un factor no especialmente relevante desde una perspectiva teórica. Minsky por ejemplo manifiesta al respecto lo que sigue:

"Para reconocer una máquina real hemos de tener alguna idea de lo que se supone que ella es capaz de hacer. El mismo conjunto de partes puede ser dispuesto, bien como una máquina de suma, o bien como una escultura moderna. Pero si se conoce cómo funciona una máquina de suma, entonces se debería reconocer una" [Min72, pp.4-5].

Por tanto, este enfoque computacional de la mente y, por ende, de la inteligencia considera en abstracto las cualidades de ciertos sistemas físicos cuya estructura es indiferente desde un punto de vista simbólico. Por esto mismo la idea de máquina involucra algún tipo de modelo o proceso abstracto. En esta línea argumentativa se podría puntualizar que, hay una importante diferencia entre la teorización en IA y la llevada a cabo en otros campos de la ciencia. Así por ejemplo, si la teoría que se toma en cuenta es la mecánica clásica (continúa señalando Minsky), y si se la considera como una generalización respecto al comportamiento de los objetos físicos, entonces, si las predicciones hechas por la teoría no se confirman (suponemos la fiabilidad de los experimentos), la teoría será revisada y quizá modificada. Algo de esto sucedió a principios de siglo con la llegada y elaboración de nuevas teorías. Sin embargo, con las máquinas y los procesos que ellas instancian, la situación se invierte por completo. Ello está relacionado con el hecho de que una máquina abstracta es una especificación de ciertos sistemas físicos que satisfacen un conjunto específico de restricciones. Así pues, ese tipo de máquina instancia la manera en la que un objeto físico determinado debería trabajar. Y la situación se invierte porque en física, cuando (en la línea del ejemplo precedente) no hay acuerdo entre la teoría y el dominio para el que debe constituir una explicación, se altera la teoría mientras que en IA, cuando la máquina, física, no funciona de acuerdo con un diseño abstracto correcto, no se modifica el diseño: se modifica la máquina. El universo no podemos cambiarlo, sólo podemos conocerlo; no pensemos que la posibilidad de sobreponernos de vez en cuando a la naturaleza supone una intervención realmente importante desde un punto de vista cósmico.

Cuando nos enfrentamos con el universo somos poco más que nada. Ello es una dificultad añadida para lograr la construcción de teorías que reflejen su funcionamiento de la forma más objetiva posible; al mismo tiempo, eso también afecta a la elaboración de buenas teorías sobre la mente y la inteligencia con las que comprender su naturaleza desde una perspectiva externa. Este es precisamente uno de los principales problemas que involucra el estudio de los procesos cognitivos. Cuando se estudia el comportamiento de un sistema físico sin embargo, en algún sentido todavía sentimos que podemos distanciarnos algo del dominio que estudiamos, en otros términos, inconscientemente se piensa que se está fuera del sistema, y ello a pesar de que la ciencia ha ido rompiendo cada vez más esa imagen. No hemos de engañarnos: estamos dentro del sistema que estudiamos. En el caso del estudio de la mente y la inteligencia, esta situación resulta aún más perplejizante, porque analizamos nuestro interior a base de reflejarlo externamente en una teoría. Es probable que esta situación de lugar a una confusión, aún mayor que en el caso de las teorías físicas por ejemplo, en lo que respecta a la elección de las variables adecuadas que mejor podrían explicar la inteligencia y sus restricciones.

BIBLIOGRAFIA

- [Fod88]: J.A. Fodor, Z.W. Pylyshyn "Connectionism and cognitive architecture: A critical analysis", *Cognition* 28, 1988 pp. 3-71.
- [Hof87]: D. Hofstadter *Gödel, Escher, Bach: un Eterno y Gracil Bucle*; Tusquets editores, 1987.
- [Min68]: M. Minsky *Semantic Information Processing*; MIT Press 1968.
- [Pen91]: R. Penrose "¿Para que sirve la mente?", *Revista de Occidente*, N° 119; Abril 1991, pp. 175-182.
- [Sim84]: G.L. Simons *Introducción a la inteligencia artificial*. Ediciones Diaz de Santos 1984.
- [Win86]: T. Winograd, F. Flores, *Understanding computers and cognition*; Ablex Publishing Corp.