

¿Pueden pensar las máquinas?

Cristian Andrés Fuentes Madariaga*

Resumen

El siguiente artículo pretende defender la posibilidad de pensamiento en las máquinas, la existencia de una tensión entre inteligencia artificial e inteligencia humana, y proponer un cierto tipo de cualidad que debería poseer una máquina para ser considerada inteligente.

Palabras clave: filosofía de la mente, inteligencia artificial, Alan Turing, androide, principio de consensualidad.

Abstract

The following article seeks to defend the possibility of thinking machines, the existence of a tension between artificial intelligence and human intelligence, and propose a certain kind of quality that should have a machine to be considered intelligent.

Keywords: philosophy of mind, artificial intelligence, Alan Turing, android, consensuality principle.

El año 1950, el lógico y matemático inglés Alan Turing propuso un test para determinar si una máquina puede pensar o no, basado en el llamado *juego de la imitación*. El juego consiste en un interrogador que realiza preguntas a un hombre y a una mujer estando cada uno de ellos en habitaciones separadas, sin poder escucharse ni verse, y cuyo medio de comunicación es sólo por escrito o a través de un mensajero. El interrogador debe realizar una serie de preguntas con el fin de determinar cuál es el hombre y cuál es la mujer, donde el hombre

* Universidad de Valparaíso, Chile.

debe encargarse de engañar al interrogador para que éste se equivoque, mientras la mujer debe tratar de entregar información para que el interrogador pueda determinar quién es quién. Turing plantea que si remplazamos al hombre por una máquina y ésta es capaz de engañar al interrogador sólo por la secuencia lógica de sus respuestas, entonces la máquina podría pensar, pues las reglas suponen necesariamente la facultad intelectual de los participantes. Pero, ¿qué tipo de máquinas son las que deben participar en este juego? Según el propósito del test, las máquinas participantes deberán ser computadoras digitales del tipo ‘máquina de Turing’, es decir, máquinas creadas para realizar cualquier tipo de operación propia de un ser humano, que puedan ser programadas para cada imitación y con una adecuada capacidad de almacenamiento, permitiendo predecir su comportamiento sin importar la cantidad de estados que pueda llegar a tener (Turing, 1950).

Si bien con este test Turing plantea alguna forma de pensamiento en las máquinas, cabe destacar que no intenta mostrar que pueda existir algún tipo de inteligencia¹ en las mismas. Pero por mero análisis del concepto, es factible sostener que la facultad de pensar supone inteligencia. Si se analiza la acepción de pensamiento como aquello que se trae a la realidad por medio de la actividad intelectual, entonces los pensamientos son productos elaborados por la mente, que pueden aparecer por procesos racionales del intelecto, o bien por abstracciones de la imaginación. Si se entiende el intelecto como potencia cognoscitiva racional de un ser humano, puede decirse que pensar es un acto racional. Entonces, si una máquina piensa, es menester que ésta tenga intelecto alguno, y si tiene intelecto, en consecuencia actúa de manera racional. Por otro lado, actuar de manera racional implica el saber elegir con el fin de asegurar nuestra sobrevivencia; por lo tanto, la facultad propia de inteligencia en un ser pensante es indispensable para su existencia. Por ende, cada próxima referencia a la posibilidad de pensamiento en las máquinas, refiere a la vez a la posibilidad de inteligencia artificial.

¹ El término inteligencia proviene del latín *intelligentia*, que a su vez deriva de *inteligere*. Esta es una palabra compuesta por otros dos términos: *intus* (“entre”) y *legere* (“escoger”). Por lo tanto, el origen etimológico del concepto de inteligencia hace referencia a quién sabe elegir.

El trabajo de Turing fue un propulsor para el desarrollo de la inteligencia artificial, ya que muchos científicos de la época dedicaron gran parte de su tiempo a fabricar una máquina que pudiese aprobar el test, como Joseph Weizenbaum, quien en 1966 creó el programa ELIZA, uno de los primeros en procesar el lenguaje natural y cuyo objetivo era mantener una conversación coherente con el usuario. Se dice que las respuestas de este programa eran tan convincentes que cuando un usuario hablaba con él, pensaba que estaba hablando realmente con un ser humano, e incluso cuando se escribía una frase que el programa no reconociera, el programa reformulaba la expresión como pregunta o reflexión.

Hoy en día las máquinas pueden sustituir a los humanos en varios aspectos laborales y domésticos. Nuestras computadoras pueden guiarnos en el desarrollo de trabajos y hasta son capaces de jugar partidos de ajedrez. Existen máquinas capaces de armar autos o realizar escáners a nuestro cuerpo para examinar nuestra salud. Todos estos inventos son fruto de la evolución de la inteligencia artificial impulsada por Turing y los científicos empeñados en superar su test.

Sin embargo, por otro lado hay un argumento que contradice el supuesto del posible pensamiento en las máquinas, propuesto por el filósofo de la mente John Searle. Se plantea que no es más que un proceso de relación de símbolos fundados en reglas encaminadas por un operador y que este proceso no incluiría la comprensión de sentidos y significados, funcionando bajo principios meramente sintácticos. Este planteamiento se basa en el experimento mental de la *habitación china*, en el que Searle se imagina encerrado en un cuarto donde se le entrega un escrito en chino (idioma que él no puede entender). Además, se le da otro escrito en chino, junto con unas instrucciones en inglés, las cuales correlacionan el primer escrito en chino con el segundo; y a continuación, se le entrega un tercer escrito de símbolos en chino junto con otras instrucciones en inglés, las cuales sirven para correlacionar el tercer manuscrito con los otros dos manuscritos entregados anteriormente. Con toda esta información, Searle es capaz de aparentar el manejo de ciertos asuntos que le son consultados, incluso puede aparentar el manejo del idioma chino entregando un nuevo manuscrito cada vez que se le presenta otro, es decir, es capaz de relacionar sím-

bolos que no conoce a través de instrucciones que sí conoce (Searle, 1983).

Por analogía, las instrucciones entregadas en inglés cumplen la misma función que cumpliría un programa de computadora. Estas instrucciones le permitirían a un ordenador entregar respuestas lógicas, coherentes y detalladas a los temas que le son consultados; en el caso de Searle, estas respuestas serían a través de un manuscrito en chino. Cabe destacar que el observador externo al cual le son entregadas las respuestas no es capaz de ver el proceso de papeleo y manipulación de símbolos que se lleva a cabo dentro de la habitación, y que además no es capaz de saber que quien da una contestación tan acertada, no tiene absolutamente ninguna idea de qué es lo que está hablando, puesto que simplemente correlacionó símbolos sin comprenderlos.

En sí, el planteamiento de Searle en cuanto a la imposibilidad de pensamiento en las máquinas refiere a su *intencionalidad*², rasgo que tradicionalmente se ha considerado como fundamental de los procesos mentales del ser humano. Según esto las manipulaciones formales de símbolos realizadas por las máquinas no poseerían intencionalidad, ya que la base de su organización sería la estructura lógica de sus respuestas y no su significado. Por lo tanto los procesos realizados por una máquina no estarían dirigidos hacia algo, es más, no serían acerca de nada (Searle, 1992).

Según Searle, los humanos realizamos manipulaciones sintácticas de signos y además conocemos su significado. Searle señala que en la medida en que conozcamos cómo funciona nuestro cerebro, tendremos una explicación más clara del problema, pues la mente existe como un aspecto del cerebro. En otras palabras, para Searle las máquinas no tienen intencionalidad y por lo tanto no piensan, mientras la intencionalidad de los seres humanos surge de su biología, característica de la que carece la composición de nuestros aspirantes a máquinas inteligentes.

En oposición a Searle, en la década de los ochenta el matemático

² La definición de intencionalidad le corresponde al filósofo alemán Franz Brentano, y consiste en el hecho de estar dirigido a algo, es decir, ser acerca de algo.

y físico estadounidense Douglas Hofstadter defendió la posibilidad de pensamiento en las máquinas con un escrito en forma de conversación (Hofstadter, 1983). En esta conversación participan tres personajes: Sandy, estudiante de filosofía, Pat, estudiante de biología, y Cris, estudiante de física. Estos estudiantes discuten entre sí dejando entrever poco a poco las ideas del autor en cuanto al pensamiento en las máquinas, que pueden exponerse como sigue: para Hofstadter, los procesos estimados tradicionalmente intencionales como pensar, desear, intentar y esperar, emergerían de la misma complejización del funcionamiento de las máquinas. Otra observación de Hofstadter señala que los computadores se han considerado como objetos totalmente fríos y cuadrados, pero si la mirada que se tiene de las máquinas cambiase, se facilitaría la idea de la inteligencia artificial, donde las máquinas podrían evocar en las mentes humanas “trazados de luz danzantes en vez de palas de vapor gigantescas”³.

La visión de un pensamiento artificial como simulación del pensamiento humano, observado en los planteamientos de Searle y Hofstadter, evidencia la tensión entre inteligencia artificial e inteligencia humana, ya que los experimentos basados en la inteligencia artificial toman como base al ser humano en vez de intentar imitar la inteligencia que poseen otros seres vivos, como perros o gusanos, que también tienen cierto tipo de inteligencia reflejada en su habilidad de aprender situaciones concretas, animales que a través del aprendizaje desarrollan la capacidad de adaptarse a un ambiente para poder sobrevivir. Otro aspecto a destacar es que cada vez que se refuta la posibilidad del pensamiento en las máquinas, en la mayoría de los casos se apela a las cualidades biológicas propias del ser humano, lo cual evidencia aún más la existencia de esta tensión entre inteligencia artificial e inteligencia humana.

Cabe mencionar que Hofstadter entiende *la simulación como equivalente a lo simulado* y, con el fin de defender la capacidad de inteligencia en las máquinas, utiliza como ejemplo la simulación com-

³ Con esto Hofstadter refiere a la posibilidad de cambiar nuestra manera de ver a las máquinas como objetos fríos y sin sentimientos, y que en la medida que cambiemos nuestro pensamiento hacia ellas, podríamos llegar a sentir un tipo de empatía, produciendo así una relación meramente emocional entre ser humano y máquina.

putarizada de un huracán, que funciona como si fuese un huracán real que modifica la programación existente dentro de una máquina, dañificando los códigos binarios unos y ceros. Según Hofstadter, en el caso del huracán simulado, si observamos la memoria de la computadora con la esperanza de ver cables rotos y destruidos, tendremos una desilusión. Pero si miramos en el nivel correcto, es decir, ya no sólo materialmente, veremos que se han roto algunos lazos abstractos que han cambiado radicalmente algunos de los valores de la programación que implican el correcto funcionamiento de la máquina. Por lo tanto, esta simulación computarizada de un huracán dentro de una máquina sería equivalente a un huracán real según Hofstadter. En sí, esta equivalencia tiene una base más bien filosófica, debido a que en este planteamiento utiliza la expresión “huracaneidad”, que representa los efectos del fenómeno y la coherencia de sus componentes, es decir, la esencia del huracán no estaría en sus componentes físicos sino en los efectos que produce. Por lo tanto, se simularía la esencia de lo simulado, infiriendo así que la esencia del pensamiento debería ser el proceso que lo subyace: el *cómputo matemático*. En consecuencia, no es de importancia el medio y el cómo se realiza la operación, sino que sólo importa la operación misma.

Este tipo de razonamiento corresponde a una corriente filosófica conocida como funcionalismo, la cual fundamenta la ciencia cognitiva presentando los fenómenos mentales en función de sus roles causales, sin depender de un constituyente físico. Dicho de otra manera, no importa si se es un computador o un humano, para el funcionalismo, la esencia del pensamiento radica en el proceso del mismo y no en su composición física. Así, el planteamiento del funcionalismo de una esencia del pensamiento que puede ser simulada por el cómputo matemático, se contrapone a la habitación china de Searle que argumenta que la mente es una característica propia del cerebro y que por lo tanto la capacidad de las máquinas de poder pensar estaría limitada por su inexistencia biológica. Gracias a esta corriente filosófica funcionalista es posible afirmar la posibilidad del pensamiento en las máquinas, el cual está ligado a la simulación del pensamiento como cómputo matemático. Y a medida de que vayamos avanzando en la comprensión de nuestra mente, también irá avanzando la posibilidad de que el cómputo matemático pueda representar la capacidad de pensamiento en las

maquinas.

Como se aprecia, la posibilidad de la inteligencia artificial está ligada a la posibilidad de imitación de la inteligencia humana, que según la corriente funcionalista puede ser representada por el cómputo matemático en sustitución del pensamiento, además cabe destacar que últimamente las máquinas a las cuales se les ha otorgado algún tipo de inteligencia tienen un aspecto físico similar al de los seres humanos, pueden representar sentimientos, y hasta detectarlos con solo mirar el rostro de una persona. En base a lo cual lo que se intenta crear no es una máquina cualquiera, sino que al parecer, lo que se trata de concebir es nada menos que un *androide*⁴, puesto que existe la intención de otorgarle a las máquinas la facultad de poseer las funciones de un ser humano para realizar acciones propias de la vida humana. En consecuencia, si se quiere avanzar mucho más en cuanto a la posibilidad de pensamiento en las máquinas y la inteligencia artificial, es fundamental preguntarnos de dónde proviene la inteligencia humana, a fin de poder definir las cualidades que debería tener una máquina para ser considerada inteligente.

Si consideramos la inteligencia como el saber elegir de manera racional con el fin de asegurar la existencia del individuo, que para el caso es nuestro androide, es necesario que éste pueda convivir de manera armoniosa con la sociedad en la que esté inserto, puesto que si no es capaz de adaptarse a la sociedad, lo más probable es que se le reprima, y por lo tanto deberá ser formateado, lo que significaría borrar todo el cómputo matemático del androide para así programar uno nuevo, más inteligente y con mayores capacidades para adaptarse a nuestro ambiente. También es necesario que nuestro androide tenga la capacidad de aprender, puesto que esta capacidad aumentaría las posibles acciones que nuestro androide pueda realizar. Además el hecho de que nuestro androide pueda aprender significaría que tendremos una máquina más eficiente, pues es posible que llegue a ser capaz de aprender hasta de sus propios errores. Estas capacidades le

⁴ Un androide es un robot u organismo sintético antropomorfo que, además de imitar la apariencia humana, emula algunos aspectos de su conducta de manera autónoma. Es un término mencionado por primera vez por Alberto Magno en 1270 y popularizado por el autor francés Auguste Villiers en su novela de 1886 *L'Ève future*.

permitirían a nuestro androide explorar de una manera más segura la sociedad en la cual este inserto, con el fin de poder desarrollar de manera autónoma su inteligencia para poder sobrevivir. Lo más probable es que existan una gran cantidad de ideas para que nuestra máquina o androide pueda ser más inteligente. Y en este sentido sería de gran ayuda recurrir a nuestra propia historia evolutiva para tener una idea de cómo nos hemos transformado en seres inteligentes.

Según Maturana, nuestra historia evolutiva es una historia de expansión de las capacidades para lograr *consensualidad*⁵, y de ahí, la expansión de inteligencia. Además, dice que la inteligencia no es primariamente la capacidad de resolver problemas, sino la capacidad de participar en la generación, expansión y operación en dominios consensuales como dominios de coordinaciones de conductas a través de la vida como individuos sociales, puesto que la solución de problemas se produce como una operación en un dominio de consensualidad establecida con anterioridad, de manera que la inteligencia es secundaria a la consensualidad, no anterior a ella (Maturana, H.1999).

Si tomamos en cuenta que vivir como humanos significa vivir en el lenguaje, el cual está compuesto de conceptos consensuados, y que como humanos es gracias al lenguaje que podemos pensar, es necesario que nuestro androide tenga la capacidad de llegar a consensos, reconociendo a los individuos con los cuales convivirá como un otro distinto a él pero legítimo. Lo cual según Maturana sería la manifestación de la cualidad de amar. Esto abre la posibilidad de una relación emocional entre ser humano y máquina, y se reduce el temor de algunos al dominio de las máquinas sobre los humanos. Este temor se ha hecho presente en varias películas de ciencia ficción, pues algunos piensan que las máquinas algún día se volverán tan inteligentes que nos verán como seres inferiores, lo cual no vendría al caso si las máquinas nos reconocieran como un otro distinto pero legítimo.

En conclusión y en base a lo que este artículo pretende defender,

⁵ La consensualidad es un principio que hace referencia al mutuo acuerdo, a la expresión de voluntad entre dos o más partes o personas. Además es un principio que rige los contratos, ya sean civiles, comerciales o laborales. Es allí donde realmente se hace evidente e importante este principio.

la facultad de pensamiento en las máquinas es posible desde una mirada funcionalista, pues esta descarta aquella inexistencia biológica que limita la capacidad de pensamiento en las máquinas dado que es algo propio de nuestra biología. Esto gracias a que según esta corriente la esencia del pensamiento no está en sus componentes físicos si no en sus efectos, los cuales pueden ser representados por el cómputo matemático. También ha quedado en evidencia la existencia de una tensión entre la inteligencia artificial y la inteligencia humana, ya que los científicos usan como base nuestra inteligencia para así programar sus máquinas, y además cada vez que se intenta refutar la posibilidad de pensamiento en las máquinas se apela a que es una facultad propiamente humana. En cuanto a la cualidad que se ansiaba proponer cabe destacar que la posibilidad de amor en las máquinas ha sido un tema bastante explorado en ciencia ficción, pero científicamente no se le ha dado el tratamiento y la importancia que merece. Por ello, a mi consideración, para juzgar como inteligente a una máquina, que además pueda aprender y relacionarse de manera armoniosa con su entorno, es menester exhortar una cualidad humanamente significativa: incluir en su cómputo matemático el reconocimiento del otro como un otro distinto pero legítimo. Si esto sucediera, lo cual es probable con el avance a pasos agigantados de la ciencia, habrá que reenfocar la preguntar: ¿es posible que un ser humano pueda aceptar a una máquina como un otro distinto pero legítimo? Esto dará paso a un fenómeno social no muy lejano, sobre la relación entre ser humano y máquina, y con ello habrá que reflexionar las futuras dificultades éticas que se podrían presentar impidiendo una armoniosa relación y los grandes beneficios a los que podríamos acceder teniendo a seres tan inteligentes como nosotros, o incluso más inteligentes, como aliados y, en lo posible, como amigos.

Referencias bibliográficas

- HOFSTADTER, D. (1983). «Temas metamágicos bizantinos. El test de Turing: conversación en un café». En D. Hofstadter y D. Dennett (comps.), *El ojo de la mente. Fantasías y reflexiones sobre el yo y el alma*. Buenos Aires: Sudamericana. Págs. 90-125.
- MATURANA, H. (1999). *Transformación en la convivencia. Desarrollo de la conservación y conciencia social e individual del niño*. Santiago de Chile: Dolmen ediciones.
- SEARLE, J. (1983). «Mentes, cerebros y programas». En D. Hofstadter y D. Dennett (comps.), *El ojo de la mente. Fantasías y reflexiones sobre el yo y el alma*. Buenos Aires: Sudamericana. Págs. 454-493.
- SEARLE, J. (1992). *Intencionalidad*. Madrid: Editorial Tecnos.
- TURING, A. (1950). «Los aparatos de computación y la inteligencia». En D. Hofstadter y D. Dennett (comps.), *El ojo de la mente. Fantasías y reflexiones sobre el yo y el alma*. Buenos Aires: Sudamericana. Págs. 69-89.