Inteligencia Humana VS Inteligencia Artificial. Un análisis critico

CIN2018A40224

Centro Universitario México

Flores García Fabián

Arturo Itzcoatl Lopez Bosquez

Humanidades y de las Artes

Filosofía

Tipo de Investigación Documental

México CDMX 16 de febrero de 2018

Tabla de contenido

RESUMEN 3
INTRODUCCIÓN 3
RESULTADOS 4
MAQUINA DE TURING 5
INTELIGENCIA ARTIFICIAL 7
LA TEORIA COMPUTACIONAL DE LA MENTE 8
REDES NEURONALES 9
CONCLUSIONES 10
BIBLIOGRAFIA 11

Inteligencia Humana VS inteligencia Artificial. Un análisis crítico

RESUMEN

Este trabajo busca exponer un análisis entre una Inteligencia Artificial (AI) y una Inteligencia Humana (HI) usando como base la Teoría Computacional de la Mente, tomando los principios del computacionalismo, y explicando cómo funcionan las máquinas de Türing, las cuales fueron cruciales para el desarrollo de la informática y que tuvieron un fuerte impacto en la creación de un modelo de la mente. Se indaga respecto de lo que es una Inteligencia Artificial, sus pros y contras, y se defiende a las redes neuronales como un modelo de aprendizaje para una Inteligencia Artificial.

INTRODUCCIÓN

PLANTEAMIENTO DEL PROBLEMA

La Inteligencia Artificial (AI por sus siglas en inglés) ha avanzado enormemente estos últimos años debido a la creciente oleada tecnológica; esto se puede apreciar, por ejemplo, en el desarrollo de software informático o en el remplazo de mano de obra humana por máquinas que aparentemente no necesitan del ser humano para operar. Sin embargo, ¿cómo podemos definir AI? ¿Es equiparable la AI con la inteligencia humana? ¿Es la AI superior a la inteligencia humana? Estas preguntas han generado en la humanidad un sentimiento de incertidumbre respecto de su futuro no sólo en ámbitos laborales o económicos, sino también en ámbitos cognoscitivos, y teleológicos, es decir, de su futuro como especie. La AI al carecer de limitaciones biológicas ¿será capaz de alcanzar o incluso superar a la Inteligencia Humana? El presente trabajo pretende exponer las teorías computacionalistas y funcionalistas, así como algunas de sus críticas de la mente para analizar críticamente este suceso contemporáneo.

HIPÓTESIS

APARENTEMENTE, LA AI AL CARECER DE LIMITACIONES BIOLÓGICAS O IMPEDIMENTOS QUE DIFICULTEN SU DESARROLLO, PRESENTARA UN AVANCE Y EVOLUCIÓN TANTO MECÁNICA COMO MENTAL MÁS RÁPIDA QUE LA DEL SER HUMANO LO QUE GENERARA SU SUPERIORIDAD SOBRE EL HUMANO.

JUSTIFICACIÓN

Uno de los temas bastante discutidos contemporáneamente dentro de la filosofía de la mente es respecto de la naturaleza de la inteligencia, de la mente, y como un tema subordinado, de la inteligencia artificial. Es indudable que delimitar el campo de la AI es crucial para prever un futuro que inminentemente tiene como base (para bien o para mal) la tecnología y la informática.

Esta investigación busca argumentos para sostener que la postura de la superioridad de la Inteligencia Artificial es superior a la Inteligencia Humana. Dando argumentos previamente recopilados en distintos textos, sobre todo de talante computacionalista. Las teorías funcionalistas y computacionalistas de la mente pueden proveer de argumentos al respecto.

METODOLOGÍA DE LA INVESTIGACIÓN

1. Búsqueda de textos que fundamenten el principio de que la AI es superior a la inteligencia humana, sobre todo en textos sobre la teoría computacionalista de la mente.
2. Análisis crítico de los textos sobre el funcionalismo mental, por ejemplo, los textos de Ned Block y David Cole.
3. Búsqueda de contraargumentos posibles a la teoría funcionalista de la mente, por ejemplo, John Searle y el argumento del cuarto chino.

RESULTADOS

Para caracterizar la Inteligencia Artificial (AI) es pertinente hacer una breve revisión histórica de las posturas que en filosofía de la mente han sido utilizadas para sostener la existencia, cualidades y desarrollo de la AI. Para ello, se revisaron diversos autores de la corriente analítica como, por ejemplo, N. Cole, G. Piccinini, N. Block, entre otros.

De acuerdo a Rescorla (CITA) Las máquinas de Türing son el antecedente principal de las teorías computacionalista y funcionalista de la mente, las cuales son utilizadas para explicar que la mente (1) puede ser entendida en términos computacionales, o que (2)

la mente puede ser entendida en términos funcionales, o bien (3) que la mente puede ser entendida en términos de un conmputacionalismo funcional[1].

Cabe acotar que las versiones más fuertes de las teorías antes mencionadas sostienen que la mente es un procesador computacional, funcional, o bien, una computadora funcional, lo que quiere decir que no sólo se explica la mente en los términos que las teorías anteriores sostienen, sino que se identifica a la mente con un procesador que computa (PC), o que realiza funciones, las cuales pueden ser o no, cálculos computacionales, en otras palabras que la mente es una computadora, sólo que en vez de tener chips y cables, tiene neuronas, tejidos, músculos, etc.

MAQUINA DE TURING

Alan Turing introdujo el concepto de "máquina de Turing" en el trabajo *On computable numbers, with an application to the Entscheidungsproblem*, publicado por la Sociedad Matemática de Londres en 1936, en el que se estudiaba la cuestión planteada por David Hilbert sobre si las matemáticas son decidibles, es decir, si hay un método definido que pueda aplicarse a cualquier sentencia matemática y que nos diga si esa sentencia es cierta o no. Turing ideó un modelo formal de computador, la máquina de Turing, y demostró que existían problemas que una máquina no podía resolver.

Con este aparato extremadamente sencillo es posible realizar cualquier cómputo que un computador digital sea capaz de realizar.Mediante este modelo teórico y el análisis de la complejidad de los algoritmos, fue posible la categorización de problemas computacionales de acuerdo a su comportamiento, apareciendo así, el conjunto de problemas denominados P y NP, cuyas soluciones pueden encontrarse en tiempo polinómico por máquinas de Turing deterministas y no deterministas, respectivamente.

Precisamente, la tesis de Church-Turing formulada por Alan Turing y Alonzo Church, de forma independiente a mediados del siglo XX caracteriza la noción informal de computabilidad con la computación mediante una máquina de Turing.

1 Las caracterizaciones de cada una de las teorías serán explicadas en los apartados consecuentes.

Una máquina de Türing es un dispositivo informático idealizado y abstracto. Alan Türing expuso por primera vez la posibilidad de crear estas máquinas en la década de los años treinta. Este dispositivo posee las siguientes características:

- Posee una capacidad de almacenamiento de información ilimitada, ya que tiene infinitas locaciones de memoria dispuestas en un orden lineal.
- Posee un tiempo ilimitado para el procesamiento de dicha información.
- El dispositivo manipula símbolos; estos símbolos son extraídos de un alfabeto finito y son guardados en ubicaciones de memoria. Por ejemplo, en el caso de un lenguaje natural como el español, las letras del abecedario serían los símbolos básicos que serían almacenados, relacionados, etc.
- Este dispositivo cuenta con un procesador central (el cual es una especie de escáner) de información el cual puede acceder a una sola locación de memoria a la vez. El procesador ejecuta cuatro acciones:
    - Escribir un símbolo en una ubicación de memoria.
    - Borrar un símbolo de una ubicación de memoria.
    - Acceder a la siguiente ubicación de memoria de manera lineal.
    - Acceder a la anterior ubicación de memoria de manera lineal.
- La tabla de máquina, la cual dicta qué operación realizará el procesador central, dado el estado del procesador y el símbolo al cual se accede. Esta tabla tiene un número finito de reglas (instrucciones mecánicas) que rigen el cálculo de la máquina de Türing. Esta tabla (se presupone) es configurada por un humano.

El procesamiento de información depende de si el símbolo está inscrito dentro de una ubicación de memoria o no, y en qué estado se encuentra el procesador de la máquina. En términos mas claros el banco de memoria es como una biblioteca; para obtener la información deseada todo depende de si se encuentra el libro (símbolo) dentro de la biblioteca (en un orden lineal) y en que parte de la biblioteca se encuentra el bibliotecario (estado en el que se encuentra el escáner o procesador central).

A pesar de su simpleza, la máquina de Türing es lo suficientemente potente como para realizar todos los procesamientos mecánicos humanamente ejecutables sobre configuraciones simbólicas, todo esto es posible debido a que imita todos los procesos

mentales básicos que genera la mente humana, como el acto de recordar o relacionar conceptos.

En el caso de una computadora personal (PC), ésta tiene limitaciones de memoria, es decir, una capacidad de memoria limitada. Se puede decir que una computadora personal es una máquina de Türing hasta que agota su memoria.

DIFERENCIAS ENTRE LA MENTE Y UNA MAQUINA DE TURING

- Una máquina de Turing funciona con símbolos puros, la mente no. (Esto quiere decir que para obtener dichos símbolos puros la mente humana tiene que depurar información previamente adquirida mediante los sentidos, por lo tanto la mente humana es menos rápida a la hora de ejecutar símbolos)
- Una máquina de Turing posee una capacidad de memoria infinita mientras que la humana no.
- La máquina de Turing ejecuta cálculos deterministas (un estado computacional determina el estado computacional posterior), la mente humana ejecuta cálculos estocásticos (el estado actual no dicta un próximo estado único si no que hay una probabilidad de que la máquina haga una transición de un estado a otro que no se encuentra directamente posterior).

INTELIGENCIA ARTIFICIAL

La pregunta central es si una máquina de Türing, o bien sus modificaciones debido a los 80 años que han pasado desde su creación, pueden desempeñar estados mentales tal cual la mente humana puede hacerlo; ¿se podrá construir una máquina capaz de pensar, es decir, de tener y realizar los mismos estados y procesos mentales que un ser humano?

El avance tecnológico posterior a la invención de las máquinas de Türing hicieron pensar a los investigadores de las ciencias informáticas que tal pregunta podría ser contestada afirmativamente, y que, en algún momento, se podría crear una máquina con capacidad de pensar, es decir, de realizar tareas mentales tales como el

razonamiento, la toma de decisiones[2] o la resolución de problemas. Ejemplos de este desarrollo tecnológico respecto de la AI sería el éxito sorprendente de IBM con Deep Blue, que derrotó al campeón de ajedrez Gary Kasparov en 1997, Watson, otro superordenador creado por IBM, logró ganar una competencia de tres días jugando al programa televisivo Jeopardy, en el que derrotó a los dos máximos campeones en la historia del programa. A diferencia de Deep Blue, Watson debió ser programada con una vasta cantidad de información, y con un motor de inteligencia artificial capaz de decidir por sí solo la respuesta más acertada a cada planteo del programa; otro ejemplo es el automóvil sin conductor Stanley, o los algoritmos de reconocimiento de voz, entre muchos otros más.

Sin embargo, una pregunta que se puede realizar en este punto es si una computadora imita el intelecto humano o verdaderamente puede reproducirlo. Para esto Alan Türing, pensó en un dispositivo de evaluación llamado “test de Türing”, el cual consiste en que una persona se interrelacione con un interlocutor incógnito, el cual es una computadora. Dado que el interlocutor permanece incógnito, la persona no sabe si es una persona con la que está interactuando o bien una máquina. Si la persona piensa que está interactuando con una persona, entonces la máquina (el interlocutor incógnito) pasa el Test. Si la persona piensa que es una máquina aquello con lo cual está interactuando, entonces la máquina no pasa el test.

LA TEORIA COMPUTACIONAL DE LA MENTE

En 1960 surgió la iniciativa interdisciplinaria *ciencia cognitiva,* que estudia a la mente recurriendo a la psicología, informativa, lingüística, filosofía, economía, antropología y neurociencia. Dicha ciencia se creó con el propósito de proporcionar un buen modelo para la mente. Para propósitos del trabajo abreviaremos Teoría Computacional Clásica

2 La toma de decisiones en ámbitos de incertidumbre resulta uno de los muchos aspectos que la AI tiene que resolver, dado que el ser humano en condiciones cotidianas realiza la toma de decisiones bajo condiciones desconocidas, o cuyo resultado no puede ser previsto o deducido de manera simple. Un ejemplo sería tener tres opciones (A, B o C) como respuestas posibles a una situación o problema específico; el ser humano escoge en algunas circunstancias, por ejemplo, la respuesta A, sin saber si A, B o C son correctas. ¿Podría una computadora realizar la misma toma de decisión bajo las circunstancias de incertidumbre especificadas? La teoría de decisiones bayesiana es el modelo matemático estándar de la toma de decisiones bajo incertidumbre, la incertidumbre se codifica a través de la probabilidad.

de la Mente en sus siglas en ingles CCTM (*classical computational theory of mind*). Para CCTM la mente es un sistema computacional [3] y similar en aspectos generales a una máquina de Türing, y los procesos mentales como el razonamiento o la toma de decisiones son cálculos similares a los que ejecuta una máquina de Türing.

Entiéndase al CCTM como un conjunto de distintas visiones respecto a la mente y no como una sola visión.

Tomando en cuenta lo que plantea CCTM la AI al igual que el humano reaccionan a estímulos, la AI recibe estímulos digitales o físicos y responde conforme a su programación, podemos decir por lo anterior que esa teoría de estímulos aplica tanto para la Inteligencia Artificial como para la inteligencia humana, un ejemplo puede ser que al usar un teclado (estímulo) la computadora genera una letra o símbolo en el monitor (respuesta), en el humano podría ser que al recibir una quemadura (estimulo) la persona genera una sensación de dolor (respuesta), por lo tanto una AI operaria de la misma manera que la de un humano ya que los dos condicionarían sus estados mentales a input y output, esto da a entender que la diferencia entre un cerebro y un CPU es meramente biológica.

REDES NEURONALES

¿Una maquina puede aprender[4]?

En la década de los 80s el conexionismo surgió como un rival al CCTM, y se enfoca en la neurofisiología, mientras que la segunda en la lógica y la informática, de esta corriente surgen las redes neuronales las cuales difieren de los modelos estilo Türing.

Una red neuronal es una serie de nodos interconectados los cuales se dividen en nodo de entrada, salida y ocultos (que funcionan como mediadores entre los dos primeros). Cada nodo tiene un valor de activación que es activado por un número real.

---

3 Esta mejor formulado referirse a la mente como un “sistema de computación” o un “sistema computacional” en lugar de una “computadora” ya que referir la mente con una “computadora” sugiere que el sistema es *programable* y esta afirmación va en contra de lo que establece CCTM de que la mente no es *programable*.

4 Entiéndase aprender como un estado mental en el cual se almacena información en algún tipo de memoria la cual puede ser utilizada posteriormente

Los pesos de una red neuronal mutan y evolucionan siguiendo un *algoritmo de aprendizaje*[5] el cual ajusta los pesos de modo que los valores se acerquen al resultado esperado.

Algunos ejemplos son la red neuronal a cargo de recomendar videos en YouTube, carros que usas redes neuronales para conducirse solos, el gran colisionador de hadrones (LHC) que utiliza redes neuronales para escoger de entre millones de colisiones las más optimas en cuestión de milisegundos, etc.

Un argumento a favor de las redes neuronales es que existe una analogía entre las redes neuronales y el cerebro ya que los nodos parecen neuronas, mientras que sus conexiones se pueden decir que tiene similitud con la sinapsis que ocurre en el cerebro humano. El modelado conexionista parece más "biológicamente plausible" que el modelo clásico.

Esto favorece en gran medida a AI ya que, a diferencia de nosotros, sus "neuronas" son digitales lo que las hace más rápidas que las humanas las cuales al ser biológicas y depender de procesos químicos son más lentas y limitadas, un ejemplo claro es que las neuronas humana mueren debido a su fragilidad mientras que las de una AI no.

CONCLUSIONES

De alguna manera podemos percibir que AI es mejor que HI debido a (como se mencionó en la hipótesis) carece de limitaciones biológicas a diferencia de la humana, también al definir los procesos mentales como procesos computacionales podemos asumir que tanto una AI y HI operan de manera similar utilizando los modelos computacionales de la mente. Ciertamente algunas desventajas que trae consigo la mente humana serían la velocidad de procesamiento y la capacidad de almacenamiento de memoria. Utilizando estos criterios podríamos concluir que AI es superior a HI.

---

5 Una manera de entenderlo más fácil puede ser tomando en cuenta al algoritmo de aprendizaje como muchísimos ejemplos que llenan de información a la red y ajustan los nodos ocultos para obtener el resultado deseado. En el caso del LHC, se llena de ejemplos de colisiones de hadrones optimas y luego la red aprende a identificar que colisiones son óptimas y posteriormente elegir en base a esos ejemplos resultados favorables

Por otro lado podemos ver que la AI requiere de intervención humana para su funcionamiento, tal es el caso de las redes neuronales al requerir la inserción de algoritmos de aprendizaje programados por humanos, al igual que el procesador central de una máquina de Turing requiere de un humano que lo opere y programe, bajo estos términos podemos concluir que AI al ser dependiente de la programación humana y HI al ser *autoprogramable*, HI es mejor. A su vez, ciertas funciones mentales parecen no ser reducibles a términos computacionales como, el imaginar o el especular.

Tomando en cuenta estas dos conclusiones se puede dar una respuesta critica: el asumir que los humanos somos *autoprogramables* puede, en cierto sentido, ser un pensamiento erróneo, ya que dicha afirmación no podemos comprobarla. ¿Cómo podemos saber que nosotros no estamos programados? Nosotros al ser los creadores de la AI podemos afirmar que son dependientes de nosotros por lo mismo que somos sus programadores. ¿Será que si en dado caso tengamos "creadores" o "programadores" y los cuestionemos respecto a nuestra "programación" solo así podamos saber si somos ajenos a cualquier tipo de programación o dependientes de una misma?

Con esta investigación se espera abrir el panorama respecto a la AI y las consecuencias que puede traer en dado caso de que se siga desarrollando la misma; también se busca dar a conocer el tema a todo el público e invitar a la reflexión del mismo, ya que se considera que es un problema bastante grande que causará conflicto en un futuro no muy lejano.

BIBLIOGRAFIA

Block, N., 1978, 'Troubles with Functionalism', in C. W. Savage (ed.), *Perception and Cognition: Issues in the Foundations of Psychology*, Minneapolis: University of Minnesota Press.

Dennett, D., 1978, 'Toward a Cognitive Theory of Consciousness', in *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, MA: MIT Press.

Cole, D. (2014). *The Chinese Room Argument en* The Stanford Encyclopedia of Philosophy. URL = <https://plato.stanford.edu/archives/win2015/entries/chinese-room/>.

Jackson, F., 1986, 'What Mary Didn't Know', *Journal of Philosophy*, LXXXIII: 291–5.

Piccinini, G. (2010). *The Mind as Neural Software? Understanding Functionalism, Computationalism, and Computational Functionalism* en Philosphy and Phenomenological Research. Vol. 81. No. 2 Missouri, United States of America: University of Missouri