INTELIGENCIA ARTIFICIAL ¿ES O NO POSIBLE?

JOSÉ JARAMILLO

25-IV-2018

Este trabajo pretende, mediante la exposición de argumentos a favor y en contra de la posibilidad de alcanzar una inteligencia artificial, mostrar que tan posible es lograr inteligencia artificial. Para comenzar es bueno realizar la distinción entre inteligencia artificial ***fuerte y débil.*** La inteligencia artificial ***débil*** es aquella que se encuentra hoy en día en la gran mayoría de la tecnología. El filósofo John Searle define inteligencia artificial débil *"La computadora digital propiamente programada es útil para estudiar la mente" y la inteligencia artificial* fuerte *"un programa de computadora literalmente tiene una mente"*[1]. La discusión principal actual está principalmente en si se puede o no lograr inteligencia artificial fuerte de forma completa. Ya que existen algunas posturas que, para argumentar que se puede lograr esta evolución en la tecnología, menoscaban o caen en un reduccionismo de la mente humana para así plantear que es posible. Además de asegurar que va a ser realizado en un cierto tiempo y con resultados muchas veces catastróficos para la humanidad. Resultados como lo son, la singularidad (unión de la raza humana con una raza más desarrollada de computadoras en una sola raza) con el fin de subsistir.

Primeramente, se van a analizar los argumentos a favor de la inteligencia artificial, después se expondrán argumentos en contra, en una postura más centrada que no habla de que existe un impedimento ontológico para lograr inteligencia artificial fuerte. Habla de que es muy poco probable que con el poco conocimiento que hay del cerebro y la mente, se logre que las computadoras verdaderamente tengan una mente. Por último, se expondrán argumentos desde una perspectiva más radical, que asegura que no se puede lograr inteligencia artificial.

Con respecto a la inteligencia artificial débil existe una dificultad bastante grande en el momento de hablar del tema. Si las computadoras sólo siguen un conjunto de reglas, para según estas mismas reglas procesar la información de entrada y lo que se pretende es una mente. Entonces se debería enmarcar la complejidad del comportamiento humano en un conjunto de reglas para que la computadora pueda procesar cualquier entrada de información. Para de esta forma poder generar comportamiento inteligente como el de los humanos. La inhabilidad de poder enmarcar todo en un conjunto de reglas lógicas es llamado el problema de la cualificación en **inteligencia Artificial**. Sobre este tema Hubert Dreyfus es uno de los principales exponentes de este argumento diciendo que los agentes lógicos son vulnerables al problema de la cualificación. A lo que el informático Stuart Russell responde que tiene la razón y que los sistemas de razonamiento probabilístico son más apropiados para los dominios abiertos como el comportamiento humano. Y es en este punto en donde Russel asegura que la crítica de Dreyfus no está dirigida a las computadoras directamente, si no una forma

[1] Mind: A Brief Introduction, John R. Searle, Oxford University Press, (Jan. 1 2005) SEP

particular tradicional de programarlas. Porque bajo la mirada de Dreyfus la experticia humana incluye el conocimiento de ciertas reglas, pero solo como "contexto holístico" o "trasfondo" en el que los humanos operan. Es en este punto donde Russel propone que Dreyfus se convierte de un crítico de la inteligencia artificial en un teórico de la misma, porque propone si darse cuenta una arquitectura neuronal, organizada en una vasta casuística. Aquí se puede notar claramente que el argumento de la informalidad no es refutado por parte de Russell, sino que es aceptado y asegura que, si es posible la inteligencia artificial, pero que el enfoque debe cambiar. Entonces no hay una refutación del argumento, hay una aceptación del argumento seguida se una intuición de que es posible hacerlo, si se intenta seguir el comportamiento del cerebro humano, cerebro que hoy en día no sabemos cómo funciona.

Según Turing para lograr una inteligencia artificial la máquina debe estar consciente de sus estados mentales y acciones. El mismo Turing también asegura que conforme las máquinas alcancen un cierto nivel de sofisticación la brecha entre inteligencia artificial fuerte y débil va a comenzar a disminuir. Para refutar la evidente cuestión fáctica de que los humanos tienen una mente real mientras que las computadoras pueden o no tener una mente real, Russell dice que *"para abordar esta cuestión se necesita entender cómo es que los humanos tienen mentes verdaderas, no solamente cuerpos que generan procesos neuro-fisiológicos"*[2]. Necesariamente entonces aparecen los esfuerzos filosóficos de intentar resolver el problema **cuerpo-mente** y son directamente relevantes para la cuestión de si las máquinas pueden o no tener mentes reales. En este punto aparecen en el debate dos posturas filosóficas con respecto al problema cuerpo-mente. Por un lado, aparece el dualismo de Descartes y por otro lado el monismo de la mente, a menudo llamado fisicalismo.

Es importante notar que los principales exponentes de este debate no toman en cuenta una postura que no sea tan extrema como puede ser la unidad sustancial de cuerpo y mente (alma), esto porque a menudo cuando se discute de inteligencia artificial es muy fácil caer en un reduccionismo de lo que es la persona humana. Un reduccionismo en lo que respecta a la complejidad de la mente, el nivel de conciencia que puede llegar a tener una persona y el gran misterio que está por detrás de toda esta complejidad unida de manera inseparable. Por lo menos hasta ahora no se ha podido separar la mente del cuerpo o viceversa, estamos lejos de entender los poderes causales del cerebro que serían los primeros indicios para comenzar a entender la complejidad de la mente humana. ¿Por qué descartar de entrada la idea de que exista en realidad una parte inmaterial? ¿Por qué tomar la posibilidad de que sea puramente material? ¿Dónde está entonces el natural deseo de trascendencia

[2] Artificial Intelligence: A Modern Approach, Stuart Russell & Peter Norvig

del hombre? Todas estas preguntas son evadidas para hablar puramente de manipulaciones eléctricas, en el caso del fisicalismo y de la posibilidad de que sea algo completamente separado del cuerpo(dualismo), haciendo una analogía rápida e imprecisa sería como un software que podría correr en cualquier hardware. Entre los desafíos más grandes que tiene los fisicalistas es explicar cómo los estados físicos, en particular, las configuraciones moleculares y electromecánicas de los procesos del cerebro, pueden al mismo tiempo ser estados mentales como sentir dolor, disfrutar una hamburguesa, saber que se está andando a caballo o más complejo aún, saber que Viena es la capital de Austria.

Si es que los fisicalistas están en lo correcto, entonces el estado mental de esa persona está determinado por el estado del cerebro de esa persona. Existe un ejemplo muy ilustrativo con el que Stuart Russell intenta explicar cómo se deben entender los estados mentales en inteligencia artificial. Es el ejemplo de un cerebro en una cubeta que consiste en remover el cerebro de una persona en el instante que termina de desarrollarse, colocarlo en una cubeta en donde se le conecta lo necesario para que crezca y se desarrolle de la misma exacta forma en que la persona hubiese vivido su vida. simulando todas las interacciones con el medio exterior, incluso la situación de comer una hamburguesa simulada. De esta forma el cerebro tendría un estado cerebral idéntico al de alguien que en realidad está comiendo una hamburguesa. Pero es falso decir que tiene un estado mental de "Saber que estoy comiendo una hamburguesa", porque no está comiendo una hamburguesa, jamás lo hizo y por ende no podría tener ese estado mental. Russell aquí hace una distinción para explicar está relación entre los estados mentales y los estados del cerebro. Explica que existen dos sentidos en los cuales se pueden interpretar los estados mentales.

El de contenido amplio[3] es el que se interpreta desde un punto de vista de un observador exterior omnisciente, con acceso a toda la situación. Que es capaz de distinguir diferencias en el mundo. Bajo este punto de vista, el contenido de los estados mentales implica el estado cerebral y la experiencia con el ambiente. El de **Contenido Angosto**[4] considera solo el estado cerebral. Donde el contenido del estado cerebral de una persona que está comiendo una hamburguesa real y el del cerebro en una cubeta es el mismo. Aquí aclara Russell que el ***contenido amplio*** es apropiado para nosotros porque nuestra meta es adjudicar nuestros estados mentales a otros con los que compartimos y para predecir cuál sería su comportamiento y sus efectos. Por eso para inteligencia artificial no tiene sentido pensar en el ***contenido amplio*** porque solo necesita el **contenido angosto**. No le interesan al sistema las

[3] Artificial Intelligence: A Modern Approach, Stuart Russell & Peter Norvig
[4] Artificial Intelligence: A Modern Approach, Stuart Russell & Peter Norvig

condiciones exteriores, lo que importa en realidad de un estado cerebral es su rol funcional en la operación mental de la entidad. Sin embargo, el ***contenido amplio*** es de gran importancia ya que existe una contraparte a la concepción cerebro-centrista como la del **contenido angosto.** Es la cognición corporizada o por su nombre en inglés "*Embodied Cognition*" que pone el énfasis en la coexistencia entre la cognición y las funciones corporales. Bajo este punto de vista las experiencias sensibles influyen directamente en la cognición. Después está la cognición embebida o por su nombre en inglés "Embeded cognition" que se refiere a la idea de que la interacción física entre la persona, el resto de la sociedad y el mundo, moldean fuertemente los posibles comportamientos de esa persona. Bajo estas miradas entonces, buscar la complejidad de la mente en la materia del cerebro sería lo mismo que buscar el valor de una moneda en el metal.

Ahora entra en juego uno de los aspectos más polémicos de inteligencia artificial, la ***conciencia***. Aparece en todos los debates de inteligencia artificial, generalmente es considerada como la capacidad de experiencia subjetiva de cualquier tipo. Desde hambre o sueño o dolor. Hasta un motivo trascendente moral o religioso. El lugar donde la persona entiende, percibe, se entiende y se percibe a sí mismo. Russell, en cambio, se concentra solo en la experiencia subjetiva, la denomina **qualia,** aclara que es un gran problema para el enfoque funcionalista de la mente porque distintas **qualias** puede estar involucradas en procesos causales isomorfos. Por ejemplo, la experiencia subjetiva de una persona x cuando observa objetos rojos, es la misma experiencia que el resto de las personas experimentan cuando observan objetos verdes. Turing dice que la cuestión de la conciencia es una cuestión complicada, pero que no tiene mucha relevancia cuando se trata de AI. Porque estos misterios no necesariamente tienen porque ser resueltos antes de que podamos responder la pregunta de si se puede o no hacer inteligencia artificial. Claramente hay una vuelta al reduccionismo, es como si fuera posible tomar la mente de partes separadas y quedarse con la parte funcional de ella, cuando en realidad la parte funcional y la parte subjetiva están unidas de formas que no entendemos todavía.

John Searle, filósofo y profesor en la Universidad de California, uno de los principales opositores de la inteligencia artificial fuerte en la actualidad, es bueno aclarar que Searle no dice que pensar en inteligencia artificial fuerte no es una cuestión imposible o ilógica. Pero que en las circunstancias y con el conocimiento que se tiene actualmente de la mente es no es posible lograr, es una cuestión fáctica. Para comenzar con la exposición de los argumentos de Searle es necesario definir la subjetividad y la objetividad, ontológicas y epistemológicas.

- Subjetividad ontológica: cuando la **existencia** de algo es subjetiva (depende de algo). La picazón, cosquillas son ontológicamente subjetivas, existen solo en la mente de la persona.
- Objetividad ontológica: cuando la **existencia** de algo es objetiva (no depende de nada). Las montañas son objetivamente ontológicas. No dependen de nada para existir.
- Subjetividad epistemológica: tiene que ver con qué tipo de **conocimiento** se tiene con respecto al objeto, si es que el saber de ese objeto depende de algún otro factor. El hecho de que Cristiano Ronaldo es mejor jugador que Messi es subjetivamente epistemológico.
- Objetividad epistemológica: tiene que ver con qué tipo de **conocimiento** se tiene con respecto al objeto, si es que el saber de ese objeto no depende de ningún otro factor. Macri es presidente de Argentina, independientemente de que algunos crean que no es apto o no ganó justamente.

Tomando este punto de partida refuta la afirmación de que la conciencia es subjetiva y la ciencia objetiva entonces no se puede estudiar por la ciencia la conciencia. Es ilógico por que la ciencia es epistemológicamente objetiva, pero tiene en cuenta los fenómenos que son ontológicamente subjetivos. Existen dos modos de existencia de la entidad según Searle, el primero objetivo e independiente y el segundo relativo y subjetivo al observador. En el caso de la computación es relativa al observador entonces no existe una ciencia de la computación por no ser independiente del observador y no tener valor intrínseco. Sino que la computación es una rama que se desprende de la ingeniería. John define la inteligencia artificial fuerte y débil, de la siguiente manera:

- Fuerte*: "un programa de computadora literalmente tiene una mente"*[5].
- Débil: *"la computadora digital propiamente programada es útil para estudiar la mente"*[6].

Aclarar estos conceptos sirve para poder exponer de manera correcta y sencilla el argumento del cuarto chino de Searle, que hasta ahora no ha logrado ser refutado por nadie. Existen intentos de hacerlo, pero como ya se expuso previamente en las refutaciones de Russell, los intentos terminan o cayendo en un reduccionismo o desviando la discusión. Este argumento se basa en qué si una entidad sigue una serie de instrucciones lógicas para llevar a cabo una determinada operación, no necesariamente entiende lo que se hizo en esa operación. Porque no tiene ningún tipo de conocimiento intrínseco de lo que está pasando. En el argumento del cuarto chino se propone que hay una persona en una habitación y recibe símbolos en chino, después de comparar con una serie de instrucciones que tiene. Devuelve los signos que debía devolver según las instrucciones. Esto no implica que la persona sepa chino, es más nunca entendió chino, porque nunca interpretó chino. Solo

[5] John Searle: "Consciousness in Artificial Intelligence"
[6] John Searle: "Consciousness in Artificial Intelligence"

intercambió y comparó signos sin entender su significado e implicancia. Y esta persona podría pasar la prueba de Turing haciendo pensar a la gente que sabe chino, cuando en realidad no entiende ni una sola palabra de chino. Un programa está definido en un aspecto puramente sintáctico y es ahí en donde tiene todo su poder, en manejar estos símbolos y esto no es suficiente porque el cerebro humano se maneja en el plano semántico. La mente humana tiene un "programa" de contenido semántico la sintaxis por sí misma no es suficiente para llegar a una semántica, no es suficiente para llegar al entendimiento***.*** *"La mente tiene semántica, los programas son puramente sintácticos, la sintaxis no es suficiente para la semántica entonces el programa no tiene una mente"*[7]***.*** Este argumento deja en claro que la inteligencia y el conocimiento intrínseco está en aquél persona que interactúa con el cuarto chino, la persona dentro del cuarto chino no tiene idea de que es el chino, podría incluso no tener conocimiento de que esos símbolos pertenecen a un lenguaje llamado chino e igual lograr los mismos resultados. Porque lo único que hace es manejarse en un set de reglas definidas sintácticamente. Según Russell en el argumento del cuarto chino, Searle depende de la intuición y no de la prueba. Sólo mira al cuarto, ¿Qué hay ahí para que sea una mente? Pero uno podría usar el mismo argumento para con el cerebro, ¿Qué hay en ese montón de células, operando ciegamente bajo las leyes de la física para que sea una mente? ¿Porque un cerebro puede ser una mente y un hígado no puede ser una mente? Ese seguirá siendo **el gran misterio**. Acá no está demás aclarar que el hecho de que sea un misterio no quiere decir que sea ilógico, o que sea un círculo cuadrado. Quiere decir que no entendemos cómo funciona, es un misterio que no sabemos si vamos a lograr entender algún día pero no es absurdo ni ilógico.

Las preguntas que plantea Russell, lejos de ser una crítica al argumento del cuarto chino, terminan estando de acuerdo porque en el cuarto chino sabemos porque no es una mente. Pero en el lado del cerebro no sabemos que hay ahí para que sea una mente y finalmente termina estando de acuerdo con Searle en que ese seguirá siendo **el gran misterio**. Es por esa pregunta que hay que comenzar y colocar todos los esfuerzos en pos de lograr responder esta pregunta y entender un poco más la mente humana. Antes de dar por sentado todo lo anterior y simplemente creer que el salto se va a lograr por alguna razón que nadie conoce.

"El sentido en el que una persona tiene información es intrínseco. El sentido en el que una persona tiene información es diferente del de las computadoras porque en las computadoras la información es relativa al observador y en la persona es intrínseca. La información que hay en una computadora está completamente en el observador y depende enteramente en la capacidad del observador de interpretar marcas en una pantalla, por ejemplo"[8]*.* Entonces esta computadora intrínsecamente tiene

[7] Mind: A Brief Introduction, John R. Searle

[8] A Discussion of Artificial Intelligence with John Searle and Luciano Floridi

solo marcas en un disco duro. La cantidad de información intrínseca en una computadora es cero. El argumento del cuarto chino señala que un programa sintáctico de computadora no contiene semántica. En ciencia cognitiva la sintaxis no es intrínseca a la física existe relativamente a la interpretación, tiene elementos de subjetividad sobre los cuáles podemos hacer aseveraciones epistemológicamente objetivas. La afirmación de que una computadora está ejecutando Word es epistemológicamente objetiva mientras que la existencia del programa en sí es ontológicamente subjetiva. Los fenómenos ontológicamente subjetivos de las computadoras no nos previenen de hacer afirmaciones epistemológicamente objetivas de esos fenómenos. El error es cuando el poder computacional obtiene su propia relevancia sicológica. Las computadoras no tienen inteligencia intrínseca. ¿Podría una máquina pensar? Nosotros somos máquinas ampliamente hablando y podemos pensar. ¿Puede una máquina hecha por el hombre pensar? no existe un obstáculo en principio para que esto ocurra. Pero no sabemos cómo el cerebro lo hace. ¿Se podría hacer con un material inorgánico? No sabemos por qué no sabemos cómo el cerebro lo hace. ¿Es la computación definida por Alan Turing suficiente para lograr que las computadoras piensen? No porque está definida sintácticamente y esto no es suficiente para la semántica.

Por último, con la ayuda del filósofo Italiano Luciano Floridi, quien es juez internacional en las pruebas de Turing que se realizan y mantiene una posición más radical con lo que respecta a la inteligencia artificial. Propone ocho argumentos comúnmente utilizados en el debate actual y los rebate, de la siguiente forma:

1. **¿Si pasa? Entonces ¿qué hacemos?**: Si pasara estaríamos en graves problemas, pero primero tiene que pasar, si el mundo estallara en 15 horas, estaríamos en un grave problema. Solo que no va a pasar, no es una afirmación completamente ilógica que el mundo estalle en 15 horas, pero ¿cuáles son las probabilidades de que pase en realidad?
2. **Podría pasar:** ¿De que podría estamos hablando? Hay dos tipos de podría, aquél podría sabiendo el futuro cercano y aquél podría que abarca cualquier probabilidad incluso la más remota, donde lo único que se toma en cuenta es que no haya contradicción. Existe una imposibilidad fáctica para conseguir inteligencia artificial, no es una imposibilidad ontológica o nomológica. El hecho de que sea una imposibilidad fáctica, significa que no es ilógico ni absurdo o que no existe ninguna regla que determine que no se puede realizar. Imposibilidad fáctica implica que existe una posibilidad, pero es parecida a la posibilidad de que Panamá sea campeón de la copa mundial de fútbol. En principio no hay impedimento ontológico o nomológico pero cualquier persona que sepa algo de fútbol sabe que no va a pasar.

3. **Crecimiento exponencial**: La realidad es que es una función sigmoide, crece y crece hasta un punto que se estabiliza. La ley de Moore, habla de que cada dos años se deberían duplicar la cantidad de transistores en un procesador. Y así se cumplió durante varios años, hasta que el mismo Moore le puso fecha de caducidad a su ley, debido a que el precio baja al igual que aumenta la capacidad de procesamiento.
4. **Tarde o temprano va a pasar:** Tardamos milenios en llegar a donde estamos y ¿de la nada va a aparecer la singularidad? Este argumento responde a teorías como las que dicen que en 2020 se da una capacidad computacional parecida a la del cerebro humano. En 2045 las computadoras superan a la raza humana. En este punto la única salida sería una "*mind uploading*" de los humanos para poder sobrevivir en una sola raza superior digital.
5. **Mejor prevenir que lamentar:** Este argumento justifica cualquier cosa no es un argumento por estar abierto a todas las posibilidades.
6. **Nos hemos equivocado antes:** Con ese argumento se puede probar cualquier cosa, es un argumento débil. Porque con un ejemplo de a y b donde en a es erróneo, si se evalúa c y d se puede asumir lo mismo que con a y b.
7. **Investigación de expertos en la materia:** Además de generar muchos ingresos siendo fanáticos de la idea lógicamente siempre dirán que es posible.

8. **Todavía no, estamos listos, ¿Sería correcto negar la posibilidad?:** ¿Existe hoy en día una remota posibilidad de mente en la computación? No, no existe. Y como no se puede asegurar cien por ciento que es imposible no significa que sea posible[9].

¿Porque las máquinas estúpidas están teniendo el éxito que tienen y no es tan difícil imaginarse que tengan verdaderamente una mente? Porque hemos transformado el mundo para estas máquinas y no vice-versa. Los códigos que leen las máquinas ya no están hechos para los ojos humanos porque no hay necesidad de que el humano lo entienda, las computadoras lo hacen automáticamente, nosotros nos adaptamos a un entorno que está hecho para las máquinas.

Para concluir, es importante mencionar los argumentos principales a favor o en contra que se expusieron. Russel habla del argumento de la informalidad donde la respuesta ante el problema de cualificación de Inteligencia artificial, es cambiar la arquitectura actual de software por una más cercana al razonamiento probabilístico. El problema cuerpo-mente las posturas a favor y en contra y la falta de estudio y análisis basados en la unidad sustancial. Los estados mentales y cerebrales, la

[9] Estos ocho argumentos son tomados de: A Discussion of Artificial Intelligence with John Searle and Luciano Floridi

conciencia y la imposibilidad de definirla completamente gracias a la falta de entendimiento actual acerca del cerebro y la mente. La objetividad y subjetividad, ontológicas y epistemológicas, los modos de existencia y el argumento del cuarto chino y la relevancia psicológica de las capacidades computacionales que expone John Searle. Por último, las refutaciones de Floridi a la mayoría de hipótesis más aceptadas comúnmente. El primer problema de los argumentos en contra es que se enfocan en evadir aquellas dificultades eminentes que tiene la inteligencia artificial, para llegar a ser verdaderamente una mente o justificar que estas dificultades son fácilmente superables. Mediante la utilización de argumentos que no son concluyentes o que prueban cualquier tesis que se pueda plantear.

Después de haber expuesto y analizado lo anterior se puede concluir fácilmente que actualmente no es posible asegurar que está próxima una inteligencia artificial fuerte, o que como dicen algunos otros más osados. La singularidad está a 10 o 15 años, porque no es una cuestión del tiempo que va a tomar alcanzar ese punto. Es una cuestión de si ¿se está buscando el lugar correcto la respuesta? ¿Cuanto esfuerzo existe en entender completamente la mente? ¿En saber cómo funciona el cerebro? Mientras se siga en la dirección actual el quiebre depende de una suerte casi milagrosa para que justo sin buscarlo se logre el mismo funcionamiento que el cerebro y sus mismos poderes causales. Lo cual es muy difícil.

Bibliografía:

- Artificial Intelligence: A Modern Approach, Stuart Russell & Peter Norvig, Prentice Hall, 3rd edition.
- John Searle: "Consciousness in Artificial Intelligence" | Talks at Google.
- A Discussion of Artificial Intelligence with John Searle and Luciano Floridi (Charla en Fritt Ord).
- Mind: A Brief Introduction, John R. Searle, Oxford University Press, (Jan. 1 2005)