EX MACHINA: INTELIGENCIA ARTIFICIAL Y NEUROBIOLOGÍA DE LOS AFECTOS

Maria Paola Caycedo Castro

Médico Cirujano Universidad El bosque, Bogota D.C

Especialista en Psiquiatria, Universidad El Bosque- Clinica Montserrat . Bogotá

Trabajo de grado para optar al título de Especialista en Bioética

Tutor: Boris Julian Pinto Bustamante

Universidad del Bosque.

Enero de 2022.

EX MACHINA: INTELIGENCIA ARTIFICIAL Y NEUROBIOLOGÍA DE LOS AFECTOS

Introducción

Ex Machina es una película de ciencia ficción de origen británico escrita y dirigida por Alex Garlan; es una película que, en tiempos de desarrollo tecnológico y ciencias cognitivas, deja más dudas que respuestas frente a uno de los grandes debates dentro de la filosofía de la mente en los últimos años: ¿Pueden las máquinas creadas por el ser humano pensar? Y más allá de eso, ¿cómo determinar si una máquina piensa y actúa de manera inteligente?.

A lo largo del artículo vinculamos esta película con algunas de las propuestas más relevantes dentro de la filosofía de la mente y la inteligencia artificial, posterior hablaremos de la neurobiología de los afectos, plantearemos como una máquina inteligente, en este caso la protagonista de la película, podría tener estados funcionales afectivos y por último intentaremos responder a un cuestionamiento que le interesa inevitablemente a la Bioética: ¿Puede un robot ser capaz de valorar?, en otras palabras ¿Tendría la capacidad de emitir juicios de valor?.

Caleb es un joven programador que trabaja en la compañía de internet más grande del mundo (Blue Book) es elegido en un concurso creado por su dueño, Nathan, para pasar una semana con él, donde tendrá que determinar si un robot humanoide con anatomía

femenina, llamada Ava, podría ser considerado humano. Todo esto basado inicialmente en el Test de Turing[1].

Durante el transcurso de la película, Caleb evidentemente empieza a dudar de su propia existencia, ya que logra ver en ese robot con estructura femenina; empatía, intencionalidad entre otras características definitorias de un ser humano.

Palabras claves: filosofía de la mente, inteligencia artificial, valores, afectos.

Filosofía De La Mente

Del Dualismo De Sustancias Al Funcionalismo.

Durante siglos, la relación mente-cuerpo ha estimulado múltiples cuestionamientos para la filosofía, la antropología, las ciencias cognitivas, las cuales confluyen en la filosofía de la mente como campo transdisciplinar, la cual pretende dar cuenta de "fenómenos mentales, buscando una explicación sistemática del mundo" (Uribe. M, 2002. p. 271). Dentro de esta corriente filosófica, hay diferentes enfoques, los cuales plantean cuestiones

[1] El Test de Turing, Fue creado por **Alan Turing (1912- 1954)** fue un matemático especialista en lógica y criptografía. Este Test consiste en que una máquina eficazmente programada mantenga una conversación con un individuo quien debe averiguar si su interlocutor es humano o máquina, solo teniendo como pista la manera como responde sus preguntas. Si este logra no diferenciarlo, se considera que la prueba es exitosa y entonces la máquina tendría Inteligencia Artificial (IA).

como: ¿Qué es mente? ¿Cómo la mente y el cerebro se interrelacionan? ¿Es posible una explicación física de lo mental?, entre otros.

A principios del siglo XVII, René Descartes abordó desde el Discurso del Método una respuesta concreta que remite a la noción del dualismo platónico que divide el mundo entre las cosas permanentes (eidos) y verdaderas (aletheia) y las cosas pasajeras o efímeras (doxa), así como a la división del mundo sensible y corruptible (el cuerpo) y el mundo de las ideas e incorruptible (alma racional). El dualismo cartesiano se entiende como dualismo de sustancia, al trazar una dicotomía entre los dos componentes de todo lo existente (a excepción de Dios): la res extensa, toda sustancia material (gobernada por las leyes de la mecánica, destructible, determinada, divisible, asequible mediante el conocimiento empírico-racionalista) y la res cogitans, la sustancia mental (indestructible, indivisible, reducto de libertad, fuente de todo conocimiento, no gobernada por las leyes de la mecánica, asequible mediante las meditaciones filosófico-religiosas).

El pensamiento cartesiano intentaba, hasta ese momento, dilucidar la relación causal entre mente y cuerpo al preguntarse: ¿cómo algo físico puede producir efecto en el alma y como eventos del alma pueden afectar el mundo físico? Descartes encuentra dicha conexión en la glándula pineal, único órgano impar, visible macroscópicamente, que se encuentra en el cerebro humano, al cual atribuyó el vínculo entre lo que sentimos, pensamos, así como el cuerpo físico y la mente (Searle, J. 2006). Este pensador, durante su vida, asumió una postura consistente de duda profunda (la duda metódica) frente a lo que percibía por sus sentidos, la cual resuelve mediante la declaración "Cogito Ergo Sum” (Pienso, por tanto, existo).

No obstante, el paradigma cartesiano ha sido replanteado de múltiples formas, en un intento por superar el dualismo de sustancias entre lo interno/externo o la dicotomía mente/cuerpo. Ludwig Wittgenstein desde la filosofía del lenguaje, aborda el problema de la mente a partir de la comprensión de los juegos del lenguaje presentes en la vida cotidiana. En Ex Machina existe una influencia de Wittgenstein, la cual es evidente, ya que Nathan llama a su gran empresa Blue Book, en honor a dicho autor, por uno de sus textos El libro Azul, cuyas notas, de 1933 a 1934, tratan sobre dichos juegos de lenguaje.

Para Wittgenstein "la filosofía de la mente se ha concentrado en el problema de la explicación científica de lo mental y por eso ha privilegiado los métodos empíricos de investigación; sin embargo, ha desatendido el problema de nuestra comprensión cotidiana de lo mental" (Pérez, M A, 2006, p. 390). Wittgenstein propone el lenguaje como una caja de herramientas cuya función, más allá de la interacción social, tiene otras funciones secundarias: intersubjetividad, pues tiene un carácter público por el entramado infinito de relaciones interpersonales que de él se derivan; así mismo tiene una finalidad que es volver visible para el otro lo que se quiere expresar, y normatividad, ya que esas interacciones sociales generan un número restringido de posibilidades para actuar. (Pérez, M A. 2006)

Además, critica el "paralelismo psicofísico", término acuñado en primera instancia por Gottfried Leibniz, según el cual, el fin último del saber científico es identificar el movimiento del pensamiento en el cerebro. Por ello, para Wittgenstein "lo que hace del pensamiento un pensamiento no es algo que se produzca en el interior del cerebro. Para que exista el pensamiento debe existir un contexto, una función para él dada por un organismo (humano)" (Moreno, K. 2011, p. 197). Se puede decir, entonces, que la "perspectiva

wittgensteiniana pretende abordar la mente en lo que de ella es relevante en la vida cotidiana de los hombres" (Pérez, M A, 2006, p. 396).

Dentro de la filosofía analítica contemporánea se puede evidenciar una vertiente muy influyente que intenta dar respuestas a las dudas que surgieron con la dicotomía cartesiana, llamada monismo, en la cual se encuentra el materialismo e idealismo principalmente. Esta última, sin tanta importancia actual por no tener un sostén científico que lo pueda respaldar, mantienen que "el universo es enteramente mental, lo que se entiende por mundo físico es solamente una de las formas adoptadas por la realidad mental subyacente" (Searl, J. 2006).

Por otro lado, los materialistas o también llamados fisicalistas sostienen que "el ser humano, incluyendo la mente, no es más que materia y complejas propiedades físicas" (Van Oudenhove,L; Cuypers, E, 2010, p. 547), y muchos de ellos terminan afirmando que lo mental se reduce a lo físico.

Algunos consideran que esta visión era un tanto problemática, ya que surgían aún más interrogantes y no daban una argumentación clara con respecto a uno de los fenómenos más importantes para los filósofos de la mente como lo es la *qualia* (entendida como las cualidades subjetivas de experiencias individuales o las propiedades fenomenológicas de los estados mentales conscientes). De lo cual se hablará con mayor detenimiento más adelante. No obstante, en nuestra película Ex machina, siempre surge la duda: ¿Ava tiene qualia, en otras palabras, tiene la capacidad de estados mentales conscientes?

Dentro de la filosofía de la mente, se buscó por todos los medios teóricos de encontrar otras vertientes para salvaguardar esta perspectiva de qualia. No obstante, aparece

en escena el materialismo eliminativo, su mayor defensor Paul Churchland, considera la inexistencia de las propiedades mentales, por lo tanto, hablar de qualia es irrelevante, así mismo no debe ser un problema la dicotomía cartesiana y se debería considerar como fuente primaria de conocimiento, no la psicología sino, las neurociencias computacionales.

Posturas algo menos radicales aparecen en escena, como sucedió en los años 50 con la Teoría de la Identidad la cual no niega la existencia de las propiedades mentales, pero postula que son ontológicamente reducible a propiedades físicas, en donde la "mente es solo cerebro y lo que imaginamos como estado mentales son solo estados cerebrales" (Searle, J. 2006).

Por último, está la posición del funcionalismo la cual considera que "las propiedades mentales no se reducen a las propiedades físicas" (Van Oudenhove,L; Cuypers, E, 2010, p. 549). En otras palabras, la postura funcionalista, en términos generales según Hierro-Pescador (2005) es:

"que los estados mentales son estados caracterizados por su posición en una cadena causal, en la cual son causa y efecto… un estado mental puede ser causa de otro estado mental o de un comportamiento y puede ser efecto de un estímulo externo o de otro estado mental… nótese que lo dicho no prejuzga que clase de entidades u objetos sean los que cumplen esa función causal." (p. 93)

Esta visión algo abstracta tiene mucha semejanza a los primeros pasos en el estudio sobre el comportamiento de ordenadores y la inteligencia artificial. El primer gran exponente de esa postura fue Hilary Putman donde toma el problema mente-cuerpo como una versión del problema de la relación entre estados lógicos y estados estructurales de una

máquina. Hierro-Pescador (2005) nos expresa lo que para Putman es que los estados lógicos de una máquina corresponden a la mente y a sus estados estructurales corresponden al cuerpo, lo interesante de esto es que Putman logra hacer un paralelismo entre estados de una máquina con los estados de un ser humano.

Del Naturalismo Biológico Al Exocerebro

Una de las corrientes con clara tendencia a reconciliar esa separación mente-cuerpo, y además es uno de los estándares más aceptados aún en la actualidad, es el dualismo de propiedad, donde parten del supuesto que "sólo la sustancia física existe. Sin embargo, éstas pueden tener propiedades tanto físicas, como mentales, que son ontológica y radicalmente diferentes entre sí, sin ser reducible lo mental a lo físico" (Van Oudenhove,L; Cuypers, E, 2010, p. 550).

A su vez, surgen de ahí dos vertientes importantes conocidas como epifenomenalismo y emergetismo; el primero ve lo mental como un subproducto de lo físico, o rasgo, en el que es totalmente dependiente y sin cabida a la autonomía; el segundo, por su parte más complejo que el primero, surge aproximadamente en los años 70 y supone un sistema complejo neurobiológico que busca dar lugar a propiedades mentales nuevas y diferentes de las propiedades físicas. Aún hoy día, muchos teóricos consideran esta perspectiva muy acorde con el modelo social y cultural del momento.

Por lo tanto, se podría deducir que el concepto de emergencia o fenómeno emergentista, no tiene un origen ontológico, sino puramente epistemológico. "Ya que las propiedades no son emergentes en un sentido absoluto, solo en relación con nuevas teorías

que pueden hacer que una propiedad que alguna vez fue emergente ya no lo sea, o viceversa" (Eronen, M. 2004, p. 21).

A finales del siglo XX y principios del siglo XXI, irrumpen nuevas propuestas teóricas más desde las ciencias computacionales en concordancia con el auge tecnológico del momento. John McCarthy entre otros, acuñaron por primera vez el término Inteligencia Artificial (IA) en los años cincuenta. Donde se llega afirmar que el cerebro es un computador digital y la mente es un programa de computador. Con la famosa analogía La mente es al cerebro lo que el programa es al hardware del computador, llegan a afirmar que cualquier sistema físico que tuviese el programa correcto con los inputs y los outputs adecuados tendría una mente (Searle, J. 2006). Los científicos dedicados a esta rama del conocimiento han llegado a afirmar que la inteligencia es solo un asunto de manipulación de símbolos físicos; por lo que si se logra la manipulación de estos se lograría "fabricar" inteligencia e incluso creencias, entre otros aspectos.

Dos de los teóricos que más han discrepado de esta propuesta son el filósofo John Searle y el físico Roger Penrose. El primero, uno de los mayores estudiosos contemporáneos de la filosofía de la mente, hace una fuerte crítica frente a este modelo. Searle nos muestra que hay vertientes dentro de la Inteligencia artificial (IA); la IA débil en donde únicamente se pretende simular estados mentales sin aspirar a que un computador tenga conciencia; y la IA fuerte, quienes, por el contrario, creen en una computadora con conciencia real y toman el Test de Turing, para consolidar su propuesta.

Esto último, es lo que considera haber logrado, sin mayores problemas, Nathan al crear a Ava, el robot humanoide, por lo que en los primeros momentos de la película se le

aclara a Caleb que él será el componente humano de la prueba de Turing. Aunque va más allá, ya que en realidad nunca se le oculta a Caleb la figura robótica con apariencia femenina porque la finalidad, según Nathan, es que a pesar de que él sepa que es un robot llegue a creer que Ava tiene consciencia.

Por otro lado, Roger Penrose en su libro magistral, La mente nueva del emperador explica cómo ha tomado gran acogida en la actualidad esta corriente y expone algunos apartes importantes respecto a la IA fuerte como la llama Searle. Para esta corriente computacional, según Roger Penrose (1996):

> "los dispositivos no sólo son inteligentes y tienen una mente, sino que al funcionamiento lógico de cualquier dispositivo computacional se le puede atribuir un cierto tipo de cualidades mentales, ya que solo consiste en una secuencia bien definida de operaciones. Para ellos lo que cuenta es simplemente el algoritmo. No hay ninguna diferencia si el algoritmo es ejecutado por un cerebro, una computadora electrónica, una nación entera de hindúes, un dispositivo mecánico de ruedas y engranajes o un sistema de tuberías. Es simplemente la estructura lógica del algoritmo lo significativo del estado mental que se supone representa, siendo completamente irrelevante la encarnación física de dicho algoritmo". (p. 27)

El Naturalismo Biológico es la culminación de años de trabajo de John Searle. El autor inicia con una fuerte crítica al manejo del lenguaje con el que categorizamos nuestra mente y cuerpo, ya que eso ha generado devastadoras consecuencias, según él, para

entender y distanciarse definitivamente de la visión tan bizarra del dualismo cartesiano. Para Searle (2006):

> "Nunca un programa de computador podrá ser una mente, ya que un programa de computador es solamente sintáctico, mientras que la mente es semántica y tienen más que solo una estructura formal; tiene un contenido... por lo tanto, tener estados mentales significa no solo manipulación de símbolos, sino tener la capacidad para interpretarlos". (Searle, J. 2006).

En Ex Machina, Caleb al tener el primer contacto con Ava, la encuentra fascinante, pero aún no está seguro de si hay inteligencia artificial real allí, por lo que posiblemente estaría de acuerdo con Jhon Searle en ese instante de la película, considerando posiblemente que Ava es únicamente un acúmulo de símbolos e información y simplemente simula tener consciencia.

Por su parte, Searle expone su crítica con el famoso experimento mental de la habitación China[2] donde refuta el test de Turing. "Los objetos biológicos (cerebros) pueden poseer "intencionalidad" y "semántica", lo que para dicho autor considera como las características definitorias de la actividad mental" (Penrose, R, 1996, p. 28).

En este punto podríamos suponer que Nathan buscaba acercarse a el concepto de Jhon Searle, ya que para él Ava inevitablemente ya paso el test de turing desde la perspectiva original, sin embargo lo que quería demostrar al llevar como elemento humano

2 Es un experimento mental, popularizado por Roger Penrose, que intenta rebatir la validez del Test de Turing a la vez que plantea que una máquina es incapaz de llegar a pensar. Expone la diferencia que existe entre reconocer la sintaxis y comprender la semántica, proponiendo que una habitación cerrada con un mecanismo dotado de la cantidad suficiente de reglas puede hacerse pasar por una persona.

a Caleb en su experimento era dejar entrever que Ava posee dicha intencionalidad y semántica de la que habla el autor.

Por otra parte, surgen otras corrientes desde una perspectiva más social, buscando una reconciliación con el medio cultural que claramente determina en ciertos aspectos nuestra mente y la concepción misma del ser humano. Autores prestigiosos dentro del estudio de la filosofía de la mente, como lo son David Chalmers y Andy Clarck, revolucionaron el concepto de mente en su texto La Mente Extendida, en donde creen que debe haber un papel activo del entorno y por tanto en la obtención de los procesos cognitivos. Por lo tanto, "no podemos simplemente señalar la piel o el cráneo como límite cognitivo para justificarnos, pues la legitimidad de ese límite está precisamente en cuestión" (Chalmers, D J. Clarck, A. 2011, p. 16). Esta visión, algo pragmática, tiene como punto central la cotidianidad de los seres humanos, ya que es allí donde se entiende que existe una dependencia profunda del entorno, donde el cerebro lleva a cabo diferentes operaciones, delegando muchas veces funciones a manipulaciones de medios externos como lo son el lenguaje, la literatura, la propia cultura, entre otros aspectos, para así extenderse fuera del sujeto.

Es claro entonces, desde este punto de vista, y nosotros estamos de acuerdo con ello, que no todos los procesos cognitivos suceden dentro de nuestro cerebro, ya que una parte del mundo funciona como un proceso cognitivo. Todo lo anterior se unifica, según los autores, en su nuevo concepto llamado externalismo activo, el cual es un sistema ensamblado, "donde el cerebro se desarrolla de un modo en el que complementa las estructuras externas, y aprende a hacer su papel dentro de un sistema unificado y densamente ensamblado" (Chalmers, D J. Clarck, A. 2011, p. 17).

La cual se caracteriza según las cuatro “E” iniciales: Extendida (extended); Embebida (embedded); Enactiva (enacted); Encarnada (embodied).

La contrariedad que surge para muchos al aceptar este concepto es que lo cognitivo excluye casi de inmediato el medio externo, y no tiene influencia sobre él. Empero, en el externalismo activo de Chalmers y Clarck, los factores externos son activos y, por lo tanto, relevantes, así que tienen un papel perentorio en el aquí y el ahora. Por lo cual tienen un impacto directo en el organismo y en su conducta.

Por último, quisiéramos detenernos en una visión desde la particularidad misma de las ciencias sociales, y de cierta manera muy concordante al externalismo activo, hallamos al antropólogo Roger Bartra con su libro Antropología del cerebro. En donde plantea como argumento central, que se deben originar, en el día a día, prótesis mentales debido a las evidentes deficiencias biológicas del cerebro para de esta manera, poder simplemente sobrevivir, sustituyendo las habilidades somáticas debilitadas. Para el autor “estas prótesis extracorpóreas no son fenómenos metafísicos, ni programas informáticos que tengan la posibilidad de separarse del cuerpo… esa prótesis es una red cultural y social de mecanismos extrasomáticos vinculados al cerebro” (Bartra, R. 2007, p. 22). Denomina esas prótesis como redes exocerebrales. En otras palabras, el autor resalta que “los circuitos exocerebrales constituyen un sistema simbólico de sustitución. Esto quiere decir que suplen ciertas funciones cerebrales mediante operaciones de carácter simbólico, con lo cual se amplían las potencialidades de los circuitos neuronales” (Bartra, R. 2007, p. 77).

Por ello, podríamos expresar que Ava es una mezcla no solo de información codificada producto de la interacción informática de Blue Book, si no además es la

terminación del contexto en donde se ha movido y ha sido creada. Por ello durante la película es claro que Ava empieza a mostrar deseos y actitudes al parecer propias de sus redes exocerebrales que implicaría su conciencia como una prótesis cultural dentro de una red simbólica, como lo expresaría Bartra. Pero ¿qué es la conciencia?

Para Jhon Searle, la conciencia es un fenómeno neurofisiológico que consiste en el conjunto de estados (procesos o eventos, etc.) mediante los cuales "sentimos y percibimos". Estos estados o procesos mentales constitutivos de la conciencia permiten que algo pueda ser considerado alguien, en virtud de quien se puede estimar el valor (o la importancia) de las cosas percibidas por los estados mentales conscientes. Estos procesos mentales permiten la relación de alguien, a través de la representación (la capacidad de representar fenómenos y objetos) y los modos de relación con el mundo sensible. Según Searle, esta relación con el mundo es posible gracias a estados mentales como: "las creencias, los deseos, esperanzas, temores, percepciones y acciones", los cuales se caracterizan por un rasgo intrínseco: la intencionalidad, la cual constituye un fenómeno de la conciencia según el cual los estados mentales tienen la propiedad de ser referidos a otros objetos distintos a sí mismo. (Searle, J. 2006).

Neurobiología y Sociología De Las Emociones/Afectos

Desde la antigüedad se ha intentado dilucidar el origen de las emociones, las cuales son parte cotidiana de todos los seres humanos, pero que a la hora de definirlas se nos hace muy difícil concretar. Por muchos años, predominó la idea de la sociología de las emociones en donde toda emoción y afecto era producto de la cultura, las instituciones, la interacción y la socialización. Se llegó a afirmar que la cultura reproduce algunas

emociones y lleva a que los individuos simplemente las expresen. Para esta corriente, poco se relaciona con el individuo y sus estados internos. Además, la gran mayoría no diferencian entre emoción y sentimiento.

Por otro lado, la neurociencia de la emoción centra el fenómeno emocional en el individuo donde se dará explicación de dichas emociones por diferentes estímulos en áreas cerebrales específicas, todo ello posiblemente como resultado evolutivo. Uno de los mayores exponentes es el neurobiólogo portugués Antonio Damasio quien traza una distinción entre las emociones y los sentimientos. Para él, las emociones son "programas de acción razonablemente complejos [...], detonados por un objeto identificable o un evento, un estímulo emocionalmente competente" (Damasio, A, 2010, p. 131). Aparecen evolutivamente como una condición de sobrevivencia, para el bienestar y equilibrio homeostático de las especies. Por ello, para Damasio las emociones las tienen también los animales. Cabe destacar aquí, que para el autor existen dos tipos de emociones las primarias y las secundarias, las cuales se explicaran más adelante. Por otro lado, los sentimientos "son procesos conscientes. La relación entre conciencia y sentimientos requiere entender que todo proceso corporal, cognitivo, emocional, supone la elaboración de imágenes producto de las redes cerebrales " (García, A. 2019, p. 51).

Se plantea una primera hipótesis donde se afirma que estos estados mentales descritos por Searle constituyen las tres dimensiones del mundo psíquico y la deliberación moral: la dimensión fáctica de la percepción (que permite la constatación sensible de las propiedades físicas de las cosas); la dimensión estimativa de los valores (cualidades afectivas atribuidas por alguien a las cosas); y la dimensión pragmática de las acciones (la intencionalidad de los actos motivados por la estimación previa de los hechos sensibles).

Para Diego Gracia, la deliberación moral exige pasar por los tres niveles, no es adecuado reducirla a una de ellas y de ahí la complejidad de esta cuestión. (Gracia. D, 2019). Por ello, es valido traer a colación el concepto de valoración de Diego Gracia:

"Valorar es una necesidad biológica tan primaria como percibir, recordar, imaginar o pensar. Nadie puede vivir sin valorar" (Gracia, D. 2010, p. 9). Entonces "la valoración es un proceso mental llevado a cabo por el ser humano en orden al logro de su objetivo biológico y vital, la supervivencia" (Gracia, D. 2010, p.10). Desde esa perspectiva los valores son emociones subjetivas que luego las transformamos en juicios de valor.

Dichos valores constan de varias propiedades: Universalidad (todos los humanos valoramos en meno o mayor medida); Pluralidad (los contenidos de los valores varían entre seres humanos); Polaridad (todos los valores tienen valores opuestos negativos o positivos); Jerarquía (Tienen diferentes niveles de importancia y esa percepción de los valores obedece a una cualidad de nuestra mente llamada estimación); Urgencia (Los valores vitales son inferiores pero fuertes, son los cimientos de los valores superiores, y por tanto los valores espirituales son superiores pero débiles); Incompatibilidad (Existe conflicto entre valores); Tragicidad (La pérdida irreparable de un valor, sobre todo de un valor intrínseco, se denomina "tragedia"). (Gracia, D. 2010). Cabria aquí volvernos a preguntar si al ver todo el transcurrir de la película Ex Machina, ¿Es posible que Ava sea capaz de valorar?

Los estados mentales, como los describe Jhon Searle, como creencias, deseos, esperanzas y temores constituyen experiencias cualitativas subjetivas (*qualia*) de orden afectivo, por lo que desde esta perspectiva, desecharíamos posturas como el materialismo eliminativo. Antonio Damasio, al final de una de sus obras, define, por ejemplo, la

esperanza como un afecto (affectus), y recurre a la definición de Spinoza: “La esperanza no es otra cosa que una alegría inconstante, que surge de la imagen de algo futuro o pasado, de cuyo resultado en cierta medida dudamos” (Damasio. A. 2007, p. 267).

Según Damasio, Spinoza emplea el término afecto como una categoría que engloba, tanto a las emociones como a los sentimientos, como modificaciones oscilantes del cuerpo y las ideas adscritas a tales modificaciones durante el proceso de ser afectado (Damasio. A. 2007). Por tanto, para Spinoza, la categoría de los afectos constituye un conjunto de “impulsos, motivaciones, emociones y sentimientos”. A partir de esta nomenclatura, Damasio propone una escala gradual de procesos regulatorios orientados a la supervivencia y el bienestar, los cuales se hacen más refinados en la medida en que progresa la complejidad evolutiva de los organismos, desde las reacciones homeostáticas básicas (metabólicas, fisicoquímicas, reflejas), pasando por los comportamientos asociados al sistema homeostático de recompensa (búsqueda y evitación, en función de percepciones de placer y dolor), el despliegue de instintos (o impulsos) y motivaciones (el apetito –como el comportamiento relativo al impulso-, y el deseo –el sentimiento consciente relativo al instinto-), como la expresión de emociones y sentimientos. (Damasio, A. 1999).

Las emociones constituyen un “conjunto complejo de respuestas químicas y neuronales que forman un patrón distintivo” (Damasio. A. 2007, p. 55) ante la presencia de estímulos emocionalmente competentes (reales o rememorados), los cuales modifican temporalmente los estados corporales y la cartografía neuronal relacionada con ellos. Las emociones son representaciones de los estados corporales cuyo objetivo es disponer tales estados en función de un comportamiento adaptativo para responder a los estímulos del entorno. Este conjunto de respuestas neuroquímicas puede adoptar distintas categorías:

emociones de fondo, emociones innatas (alegría, tristeza, repugnancia, ira, sorpresa, temor) y emociones sociales, mediadas por el aprendizaje social (vergüenza, culpa, indignación, pudor, etc.).

Entonces es claro que las emociones primarias son innatas “dependen de la circuitería del sistema límbico, siendo sus principales actores la amígdala y la corteza cingular anterior” (Damasio, A. 1999, p. 157). Por otro lado, aparecen las emociones secundarias, que se vinculan inicialmente con la amígdala por lo que utilizan la maquinaria de las emociones primarias, pero “es analizado por el proceso pensante y puede activar las capas corticales frontales” (Damasio, A. 1999, p. 162), entre otras estructuras cerebrales.

Por otro lado, los sentimientos se configuran entonces a partir de una idea (o un conjunto de ideas) asociadas a la modificación emocional de los estados corporales de manera consciente. Para Damasio, las ideas (y los pensamientos) son imágenes mentales derivadas de un conjunto de patrones de actividad neuronal (cartografías sensoriales) que permiten la representación de los estados corporales en relación con objetos del mundo externo que interactúan con el organismo, y que cumplen diversas funciones reguladoras. En tal sentido, las ideas son categorías análogas a los estados mentales intencionales propuestos por Searle.

Podemos refinar un poco esta descripción, y aquí plantear una segunda hipótesis, afirmando que una idea es un estado mental intencional, por medio del cual la conciencia representa el mundo y despliega las condiciones de posibilidad para relacionarse con él. En tal sentido, la consciencia no es solo un atributo autor referencial, pues como afirma Roger Bartra:

> "Podemos entender la conciencia como una serie de actos humanos individuales en el contexto de un foro social y que implican una relación de reconocimiento y apropiación de hechos e ideas de las cuales el yo es responsable. La manera en que Locke ve a la conciencia se acerca más a las raíces etimológicas de la palabra: conciencia quiere decir conocer con otros. Se trata de un conocimiento compartido socialmente. En este sentido, las ideas no constituyen simples representaciones mentales, sino "unidades de trasmisión de información cultural" símbolos y motivos que, asociados a los estados emocionales, configuran los sentimientos" (Bartra, R. (3 de septiembre de 2014).

En este punto nos separamos de la noción que defiende Damasio en cuanto a la naturaleza privada de los sentimientos. Si bien concordamos en que las emociones se despliegan en el teatro del cuerpo, mientras los sentimientos lo hacen en el teatro de la mente, no concordamos con la afirmación que remite los sentimientos al ámbito privado de la conciencia, pues las percepciones y estimaciones de la vida emocional conservan la posibilidad de ser comunicados y son, a su vez, un producto de la comunicación. Su papel no se reduce a la dimensión de razonabilidad asociada a los mecanismos emocionales automáticos para la gestión de problemas complejos. Los sentimientos, como nociones comunicables, son a la vez fundamento y producto de la red simbólica que constituye la cultura, como conciencia colectiva de sentimientos compartidos. La noción de emociones sociales, como la vergüenza, el pudor, la culpa, y su relación con emociones primarias y emociones de fondo, sólo pueden ser comprendidas en función de la existencia de símbolos culturales de carácter normativo (nociones como el pecado, el castigo, la redención, etc.),

presentes en circuitos exocerebrales de significado social. Al referirse John Searle a la dimensión semántica del lenguaje, entendida como la atribución de significado a la sintaxis de los símbolos lógico-formales, cabe precisar, desde una perspectiva antropológica, que la sintaxis, como los significados, "se construyen en una red que conecta circuitos neuronales con redes culturales" (Bartra, R. 2007, p. 129).

En este punto es posible, como una tercera hipótesis, vincular afectos y sistemas simbólico-culturales (como el lenguaje, los ritos, la música, la danza, los mitos, etc.). Las emociones sociales y los sentimientos se configuran a partir de un proceso continuo de retroalimentación entre señales neuronales y símbolos culturales, en lo que Bartra ha denominado el exocerebro, como "un circuito extrasomático de carácter simbólico" (Bartra, R. 2007, p. 25) que sustituye funciones neuronales deficientes (olfato, audición, visión nocturna, etc.) en términos adaptativos, mediante el desarrollo de estructuras cerebrales generadoras de sistemas de codificación lingüística y simbólica (áreas de Broca y Wernicke), relevantes para una especie animal en condiciones de precariedad biológica y necesitada de la prótesis cultural de los intercambios simbólicos, sobre los cuales se erige la cultura, como una segunda naturaleza artificial.

Desde esta perspectiva, las emociones y los sentimientos, como fenómenos neuroculturales, constituyen estados mentales intencionales que cumplen cuatro propiedades: son subjetivos (son experimentados por el sujeto, en primera persona, y son intransferibles), internos (la interacción entre señales endocerebrales y símbolos o pautas culturales se codifica en patrones neuronales y estados corporales), cualitativos (se experimentan como sensaciones o sentimientos valorativos) y unificados (estas representaciones se presentan como procesos coherentes).

En este punto es importante nuevamente traer a colación la noción de los qualia, como experiencias subjetivas cualitativas de cualquier tipo generadas por el sistema nervioso, sean estos sentimientos o sensaciones. La analogía de la habitación de Mary[3], intenta describir la relevancia de estas qualias para la noción de la consciencia.

Estos estados mentales subjetivos, internos, valorativos e integrativos constituyen, como expresamos anteriormente, afectos. La dimensión semántica a la que remite la propuesta de Searle, como atribuciones de significado a las configuraciones meramente sintácticas del lenguaje, no puede comprenderse sin el intercambio cultural de los afectos (desde los impulsos y las motivaciones, hasta las emociones y los sentimientos). En este punto, la relevancia, del cuerpo, el movimiento, la sexualidad y los significados atribuidos a partir de los circuitos simbólicos de la cultura, son dimensiones relevantes, tanto para la percepción, como para la estimación de los fenómenos del mundo sensible.

3 Es un experimento mental que plantea Frank Jackson como crítica al fisicalismo. Mary es una científica brillante que investiga el mundo desde un cuarto blanco y negro a través del monitor de una televisión en blanco y negro. Se especializa en la neurofisiología de la visión y adquiere, toda la información física que hay para obtener acerca de lo que sucede cuando vemos tomates maduros, o el cielo, y usa términos como «rojo», «azul», etc. Pero nunca ha experimentado el color. Entonces el autor se pregunta ¿Qué sucederá cuando Mary sea liberada de su cuarto blanco y negro o se le dé una televisión con monitor en color? ¿Aprenderá algo o no?

Conclusiones

En la película es posible afirmar que, si bien Ava no ha interactuado con otros seres, más allá de Nathan y Caleb, su wetware está dotado de un universo de información obtenida a partir de los hábitos de navegación y las preferencias de consumo de todos los usuarios de Blue Book (los cuales constituyen, al tiempo, estados mentales afectivos), configurando un sistema robótico dotado de una especie de intencionalidad algorítmica, la cual, a su vez, interactúa con un sujeto intencional quien también está dotado de un wetware (cerebro) que integra un universo de información biológica y cultural (y que incluye atributos genéticos, sensores epigenéticos, pautas culturales de socialización, etc.). El rostro de Ava, por ejemplo, ha sido diseñado por Nathan a partir de los hábitos de consumo de pornografía de Caleb, en lo cual pueden converger, tanto atributos hereditarios vinculados con la sexualidad y la reproducción humana, como información simbólica mediada por los estereotipos culturales del deseo, en una especie de algoritmos neurobiológicos.

Entonces, el test de inteligencia artificial fuerte expresado en *Ex machina* exige responder las siguientes preguntas: ¿La robot es intrínsecamente intencional? ¿Experimenta estados mentales subjetivos y cualitativos? ¿Es capaz de valorar? ¿Es posible que logre estimar los fenómenos más allá de la percepción? ¿Esas estimaciones sugieren una convergencia entre algoritmos (equivalentes a las señales de información de los sistemas neurobiológicos) y los símbolos culturales? ¿Esas valoraciones cualitativas constituyen estados mentales afectivos (esperanza, temor, deseos, creencias)?

Según nuestra opinión, la respuesta a estas preguntas, según lo expuesto en la película, es afirmativa. Los estados mentales intencionales propuestos por Searle se pueden concentrar en las tres categorías de afectos propuestas por Damasio: (a) Impulsos o motivaciones: apetitos y deseos (b) Emociones: esperanzas y temores (c) Sentimientos: creencias.

Por tanto, la hipótesis final defiende la naturaleza afectiva, cualitativa y subjetiva de tales estados mentales, así como su dimensión simbólica mediada por la cultura. En este sentido, la prueba de la inteligencia artificial fuerte consiste en la constatación de estados mentales afectivos, valorativos mediados por la interacción cultural. Estos estados mentales exigen una configuración material o corpórea (no necesariamente orgánica) que permita al sistema la representación y valoración de los estados internos de configuración afectados por la interacción con estímulos emocionalmente competentes provenientes del entorno físico y simbólico. Si un sistema artificial está dotado de los repertorios sensoriales e integrativos, capaces de incorporar no solo la información sintáctica proveniente de estímulos externos, sino también las pautas culturales, a partir de las cuales sea posible la generación de estados funcionales afectivos (emocionales, sociales y sentimientos), tal sistema artificial habrá superado definitivamente el test de Turing.

En otras palabras, consideramos que esta película de ciencia ficción propone lo anteriormente expuesto como un desafío a la Inteligencia Artificial, ya que si definitivamente se quiere crear una prueba para determinar que la IA posee consciencia de sí misma, creemos que no es otra cosa que que lo que logra Ava en Ex Machina, estados mentales afectivos.

Bibliografía

Uribe, M. (2002). Epistemología, filosofía de la mente y bioética. Rev. Colombiana de psiquiatría, 31 (4);271-4. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0034-74502002000400007

Searle, J R. (2006). LA MENTE una breve introducción. Grupo editorial norma.

Pérez, M A. (2006). Mente y relevancia. Univ. Psychol, 5 (2). 385-96. http://www.scielo.org.co/pdf/rups/v5n2/v5n2a14.pdf

Moreno, K. (2011). Wittgenstein y la naturalización de la mente. Tesis psicológica, (6). 183-200. https://www.redalyc.org/pdf/1390/139022629012.pdf

Van Oudenhove,L. Y Cuypers, E. (2010). The Philosophical "Mind- Body Problem" and Its Relevance for the Relationship Between Psychiatry and the Neurosciences. *Perspectives in Biology and Medicine*, 53 (4). 545-557. doi:10.1353/pbm.2010.0012.

Hierro-Pescador, J. (2005). Filosofía de la mente y de la Ciencia cognitiva, Editorial AKAL.

Eronen, M. (2004). Emergence in the philosophy of mind. Helsinki (Finlandia). 1-83.

Penrose, J. (1996). La mente nueva del emperador: En torno a la cibernética, la mente y las leyes de la física. Fondo de cultura económica.

Chalmers, D J. y Clark. A. (2011). La mente extendida. Cuadernos de información y Comunicación, (16). 15-28. https://revistas.ucm.es/index.php/CIYC/article/view/36985/35794

Bartra, R. (2007). Antropología del cerebro, la conciencia y los sistemas simbólicos. Fondo de cultura económica.

Damasio. A. (2007). En busca de Spinoza. Neurobiología de la emoción y los sentimientos. Editorial Drakontos.

Bartra, R. (3 de septiembre de 2014) Uno de los más grandes enigmas). El Tiempo. https://www.eltiempo.com/archivo/documento/CMS-14478635

Damasio A (2010). *Self Comes to Mind.* Nueva York: Vintage Books.

Garcia, A. (2019). Neurociencia de las emociones: la sociedad vista desde el individuo. Una aproximación a la vinculación sociología-neurociencia. Sociológica, 34 (96). 39-71. http://www.scielo.org.mx/scielo.php?pid=S0187-01732019000100039&script=sci_abstract

Damasio, A. (1999). The Feeling of what happens. Body and emotions in the makingof consciuness. The New York Times Book Review Editors.

Damasio, A, 1999. El error de Descartes. La razón de las emociones. Editorial Andres Bello.

Gracia, D, 2019. Bioética Mínima. Editorial Triacastela.

Gracia, D. 2010. La cuestión del valor. Real Academia de Ciencias Morales y Políticas