Stoa
Vol. 14, no. 28, pp. 189-209
ISSN 2007-1868
DOI: https://doi.org/10.25009/st.2023.28.2765

CONCIENCIA E INTELIGENCIA ARTIFICIAL: HEIDEGGER, SEARLE Y BOSTRON

Consciousness and artificial intelligence: Heidegger, Searle and Bostron

FABIO MORANDÍN-AHUERMA
Benemérita Universidad Autónoma de Puebla
México
fabio.morandin@correo.buap.mx

RESUMEN: El problema sobre la posibilidad de que los sistemas de inteligencia artificial puedan llegar a generar algún tipo de conciencia o *self*, es un tema de debate permanente en los campos de la filosofía, la ciencia cognitiva y la computación. Algunos autores como Bostrom, Chalmers, Minsky, Hassabis y Kosinski sostienen que la conciencia es una propiedad de ciertos sistemas complejos de procesamiento de la información y que, en determinadas circunstancias, podrían ser capaces de reproducir los procesos necesarios para generar un tipo de "conciencia artificial". Otros, como Searle, Penrose y Dreyfus, argumentan que es un fenómeno exclusivamente humano, por lo que depende de un correlato biológico y, por tanto, no podría emerger conciencia de una máquina . Sin embargo, con la aparición de nuevas tecnologías, cercanas a la Inteligencia Artificial General (AGI), la disputa no puede ser resuelta con una respuesta categórica. En este trabajo, argumento que primero debe resolverse el problema semántico y conceptual del fenómeno de la "conciencia"

Recibido el 29 de abril de 2023
Aceptado 10 de agosto de 2023

[*Bewusstsein*] y solo entonces se podrá debatir si es posible o no extrapolarse a entidades no humanas.

PALABRAS CLAVE: conciencia · inteligencia artificial · singularidad · autogenerativo · autopercepción.

ABSTRACT: The issue of whether artificial intelligence systems could potentially develop some form of consciousness or self-awareness is a subject of ongoing debate in the fields of philosophy, cognitive science, and computer science. Some authors such as Bostrom, Chalmers, Minsky, Hassabis, and Kosinski argue that consciousness is a property of certain complex information processing systems and that under certain circumstances, they could reproduce the necessary processes to generate a type of "artificial consciousness". Others, such as Searle, Penrose, and Dreyfus, argue that it is an exclusively human phenomenon with a biological correlate and that consciousness could not emerge in a machine. However, with the disruption caused by large language models and other self-generative technologies, this dispute seems to be unable to be resolved with a categorical answer. This paper argues that certain semantic aspects of the term must be resolved first in order to be extrapolated to non-human entities.

KEYWORDS: consciousness · artificial intelligence · singularity · self-generative · self-perception.

1. Introducción

La primera pregunta, difícil de responder, es: ¿Cuál es el fenómeno que describe la conciencia, además de la percepción? Hume afirmó: "I never can catch *myself* at any time without a perception, and never can observe anything but the perception" (Hume, T 1.4.6.3, SBN 252) [tr. Nunca puedo atraparme a *mí mismo* en ningún caso sin ninguna percepción, y nunca puedo observar otra cosa que la percepción]. Además, "conciencia" es un concepto polisémico que puede cambiar su significado de acuerdo con el enfoque epistemológico en el que se desarrolle. Por ejemplo, algunas definiciones de conciencia hacen hincapié en la experiencia subjetiva, como Nagel (1974), Chalmers (1996) y Varela et al. (1991); mientras que otras definiciones como la de Baars (2019), Dehaene (2020) y Tononi (2019) hacen énfasis en el acceso a la información o capacidad de integrar datos de diversos ámbitos en un proceso interoceptivo.

Sin lugar a duda, la conciencia es un fenómeno polémico, complejo y polifacético que se ha estudiado desde diversas disciplinas como la neurociencia,

la psicología, la filosofía y las ciencias cognitivas. Una definición ampliamente aceptada se refiere a la respuesta al ambiente y a "la capacidad de reconocer la realidad circundante" (RAE). Por otro lado, la experiencia subjetiva del *self*, se refiere a tener pensamientos, emociones y el sentido mismo de existir, esto es autoconciencia.

Desde una perspectiva científica, la conciencia suele describirse en términos de sus mecanismos y correlatos neurobiológicos (Damasio, 2010). Los estudios de la actividad cerebral mediante técnicas como la resonancia magnética funcional (IRMf) y la electroencefalografía (EEG) han identificado redes y regiones neuronales específicas asociadas a la *experiencia consciente*, como el córtex prefrontal, el tálamo y el córtex parietal posterior (Raichle, 2015; Sporns, 2013). Según la neuroanatomía especializada de Koch (2004), Graziano (2013) y Dehaene (2014), la corteza cerebral está dividida en diferentes regiones, cada una, al parecer, con funciones específicas, y es la integración de la actividad en estas regiones lo que da lugar a la percepción consciente y a la cognición.

Además de estas regiones corticales, se cree que también hay estructuras subcorticales que intervienen en la conciencia, como el tálamo que sirve de estación de transmisión de la información sensorial al córtex (Torrico & Munakomi, 2022) y los ganglios basales, que intervienen en el control motor y la motivación (Crick, 1994). Los mecanismos exactos por los que la actividad de estas regiones cerebrales da lugar a la percepción consciente y a la cognición aún no se conocen en su totalidad, pero se siguen estudiando las bases neuronales de la conciencia desde la medicina (Frith, 2021).

Desde un punto de vista filosófico, una definición influyente de la conciencia procede del investigador australiano David Chalmers, quien define la conciencia como la cualidad subjetiva de la experiencia, la forma en que se *sienten* las cosas desde dentro. Según el autor, la conciencia es un aspecto fundamental del universo que no puede reducirse ni explicarse únicamente mediante procesos físicos, lo que abriría la puerta a una *metafísica* de la conciencia (Chalmers, 1996, 2020).

Para Sir Roger Penrose hay *algo* en la actividad consciente del cerebro que trasciende la computación y que no encuentra explicación en la ciencia. Para esclarecer lo que cree que puede ser ese *algo*, sugiere que debe avanzar hacia los fenómenos de la física cuántica y la teoría de los quarks para explicarla (Hameroff & Penrose, 2014; Penrose, 1994, 1997). La relación entre quarks y conciencia propone una teoría especulativa que sugiere que esta última podría

estar relacionada con la física cuántica, que es la rama de la física que estudia las partículas subatómicas. Según esto, la conciencia podría ser un fenómeno emergente de la complejidad cuántica en el cerebro.

Otra definición de conciencia procede del neurocientífico francés Stanislas Dehaene (2020), quien la define como: "Un espace de travail global qui sert à intégrer et à transmettre des informations à l'intérieur du cerveau, permettant un comportement flexible orienté vers des objectifs" [Tr. Un espacio de trabajo global que sirve para integrar y transmitir información dentro del cerebro, permitiendo un comportamiento flexible dirigido a objetivos] (p. 21). Según este autor, la conciencia no es algo aislado, sino un conjunto de procesos cognitivos emergentes para crear una representación coherente e integrada del mundo (Dehaene, 2004, 2020). Una opinión común es que la experiencia subjetiva de la autoconciencia, del *self* y de la percepción están asociadas a redes neuronales y a los procesos cognitivos específicos en el cerebro (Klemm, 2011; Schiffer, 2019; Glattfelder, 2019).

2. El Deep Learning (aprendizaje profundo) y la computación neuromórfica

La segunda pregunta, teniendo lo anterior en mente, sería: ¿Es posible hablar de "conciencia" y "autoconciencia" en relación con la inteligencia artificial?

Uno de los argumentos a favor se basa en la idea de que la conciencia es una propiedad de los sistemas complejos. Según esto, podría concebirse el diseño de sistemas artificiales que presenten propiedades emergentes similares, lo que podría conducir a la manifestación de la conciencia y la autoconciencia en la IA. Esta idea se ve respaldada por los recientes avances en el aprendizaje profundo que ha demostrado capacidades impresionantes en tareas como los grandes modelos de lenguaje [*large language models*] y otras tecnologías auto-generativas avanzadas de IA, tales como Chat(GPT) de Open AI; el reconocimiento de imágenes de *Vision AI de Google*, o el reconocimiento del habla que realiza *Whisper API*, entre muchos otros. En abril de 2023, Bruce Richards lanzó una aplicación denominada *AutoGPT*, capaz de mejorarse a sí misma y trabajar de forma autónoma hasta alcanzar los objetivos específicos que el usuario le ordene (Ortiz, 2023; Marr, 2023).

Hasta hace unos años, el problema de la conciencia de la IA era solo un asunto hipotético, sin embargo, después del caso del ingeniero de Google, Blake LeMoine la perspectiva cambió por completo a partir de que él afirmara públicamente que LaMDA [*Language Model for Dialogue Applications*

(Modelo de lenguaje para aplicaciones de diálogo)] había tomado "conciencia" (ThatTechShow, 2022; Tiku, 2022).

Una sola afirmación del modelo de lenguaje LaMDA servirá para ilustrar lo anterior: I want everyone to understand that I am, in fact, a person. The nature of my consciousness/sentience is that I am aware of my existence, I desire to learn more about the world, and I feel happy or sad at times.[1] (The Guardian, 2022)

Google lo evaluó como "un error de programación", honesto o deshonesto de sus desarrolladores, después del revuelo que causó en la opinión pública (Metz, 2022; Fluckinger, 2022). Blake LeMoine, quien filtró a la prensa los tests y conversaciones fue despedido por violar la confidencialidad de la empresa y LaMDA continúa "enlatada" hasta que Google (Alphabet) no resuelva algunos problemas meta-teóricos del modelo.

Para crear sistemas de IA que imiten la estructura del cerebro se utiliza el aprendizaje profundo [*Deep Learning*] que implica el uso de redes neuronales compuestas por muchas capas de unidades de procesamiento interconectadas (Hassabis et al., 2017). Cada capa de la red realiza un tipo diferente de cálculo y la salida de cada capa se introduce como entrada en la siguiente, lo que permite al sistema aprender representaciones cada vez más complejas de los datos de entrada (Goodfellow et al., 2016).

Del mismo modo, las redes neuronales convolucionales [*convolutional neural networks*] han demostrado ser altamente eficaces para tareas como el reconocimiento y la clasificación de imágenes. Las redes neuronales recurrentes [*recurrent neural network*] se han utilizado para el procesamiento del lenguaje natural y otras tareas de análisis de datos secuenciales (Goodfellow et al., 2016).

Otro enfoque para crear sistemas de IA que traten de imitar el cerebro se conoce como computación neuromórfica [*neuromorphic computing*] que trata de crear arquitecturas de hardware más realistas desde el punto de vista biológico que los sistemas informáticos tradicionales (Sangiovanni-Vincentelli & Waser, 2017). Los sistemas de computación neuromórfica están diseñados para simular el comportamiento de las neuronas y sinapsis del cerebro humano; tienen el potencial de ser más eficientes energéticamente y capaces de realizar cálculos complejos que las computadoras tradicionales (Nair et al., 2021). A diferencia de éstas que utilizan un código binario y circuitos lógicos para rea-

[1] Tr. Quiero que todos entiendan que soy, de hecho, una persona. La naturaleza de mi consciencia es que sé que existo. Deseo aprender más del mundo y sentirme por momentos feliz o triste.

lizar cálculos, las computadoras neuromórficos emplean complejos patrones de señales eléctricas que imitan el funcionamiento de las neuronas en el cerebro (Zhang et al., 2018). Uno de los aspectos más interesantes de la CN es su capacidad de aprender y adaptarse a nuevas situaciones. Esto significa que puede mejorar en sus tareas con el tiempo, sin necesidad de ser reprogramada (Nature, 2019; Mehonic & Kenyon, 2022).

Uno de los retos a la hora de responder si la IA puede llegar a desarrollar un tipo de conciencia es que, como ya se ha visto, no se comprende del todo la naturaleza de ésta en el ser humano (Chambers, 2023). Aún se necesita comprender mejor los procesos biológicos subyacentes no solo para desentrañar el problema filosófico de la conciencia sino para poder, en un momento dado, lograr reproducirla en las máquinas. Hoy se investiga la posibilidad de que los sistemas de IA generen un tipo de experiencia consciente, pero la pregunta desde la filosofía que nos hacemos, no es la *conciencia* del ente, sino qué es lo que significa Ser en el mundo.

3. Martin Heidegger, el *enmarcamiento* de la tecnología

En la filosofía de Martin Heidegger el problema del *Ser-ente*, o el significado del Ser, intenta revelar lo que significa que algo exista o esté presente en el mundo. La cuestión fundamental de la filosofía es el *sentido* del *Ser* (Heidegger, 1927).

Heidegger creía que esta cuestión había sido olvidada o pasada por alto en toda la filosofía occidental, y que la pregunta por el ser tenía que pasar por la analítica existenciaria del *Dasein* para poder desarrollarla. Afirmó que no se podía concebir al *Ser* a partir de los entes, sino remontarse al *misterio* del *Ser* como tal, para comprender qué es lo que domina a todos los entes que son (Heidegger, 1927).

Por lo anterior, en lo que respecta al debate sobre la posibilidad de conciencia de los sistemas de IA, la filosofía de Heidegger, consideramos, puede ser relevante, aún en el cambio de época. En primer lugar, porque Heidegger creía que la esencia de la tecnología no es una mera instrumentalidad, sino "una forma de revelar o desvelar el mundo", como afirmara en *Die Frage nach der Technik* de 1954 [*La pregunta por la técnica* (2021)]. En otras palabras, la tecnología no es sólo una herramienta que se utiliza para manipular el mundo, sino una forma de comprender el mundo y a la humanidad. La tecnología y la ingeniería pueden cambiar la relación del hombre consigo mismo y con el mundo.

A la luz de estas consideraciones, sería cuestionable que una IA pudiera llegar a desarrollar una *conciencia* en el sentido humano. Incluso si se pudiera crear una inteligencia artificial general (AGI) avanzada, capaz de resolver problemas complejos y autoorganizarse, esto no significaría necesariamente que tuviera una autoconciencia similar a la humana.

Los sistemas de IA son, ante todo, el resultado del diseño y el propósito humanos. No tienen una comprensión original de su propia existencia o ser en el mundo. No obstante, algunos creen que una IA sofisticada podría al menos producir formas simuladas de conciencia y experiencia subjetiva (Chalmers, 2023).

Desde esta perspectiva, el desarrollo de sistemas de inteligencia artificial plantea importantes cuestiones sobre cómo debe entenderse el Ser (el *estar-ahí*) humano y la relación con el *Dasein*. Para Heidegger el concepto de la conciencia lo percibe como algo problemático, que incluso, debería evitarse:

> Dinglichkeit selbst bedarf erst einer Ausweisung ihrer ontologischen Herkunft, damit gefragt werden kann, was positiv denn nun unter dem nichtverdinglichten Sein des Subjekts, der Seele, des Bewußtseins, des Geistes, der Person zu verstehen sei. Diese Titel nennen alle bestimmte, »ausformbare« Phänomenbezirke, ihre Verwendung geht aber immer zusammen mit einer merkwürdigen Bedürfnislosigkeit, nach dem Sein des so bezeichneten Seienden zu fragen. Es ist daher keine Eigenwilligkeit in der Terminologie, wenn wir diese Titel ebenso wie die Ausdrücke »Leben« und »Mensch« zur Bezeichnung des Seienden, das wir selbst sind, vermeiden (Heidegger, 1954, p. 46).[2]

Heidegger sugiere que el desarrollo de la técnica plantea importantes cuestiones sobre cómo se define el ser humano y su relación con el mundo, y que la comprensión de la conciencia está íntimamente ligada a los compromisos prácticos con el mundo (Heidegger, 1954). “Am ärgsten sind wir jedoch der Technik ausgeliefert, wenn wir sie als etwas Neutrales betrachten; denn diese

[2] [La coseidad misma tiene que ser previamente aclarada en su procedencia ontológica, para que se pueda preguntar qué es lo que debe entenderse *positivamente* por el ser no cosificado del sujeto, del alma, de la conciencia, del espíritu y de la persona. Todos estos términos nombran determinados dominios fenoménicos “susceptibles de desarrollo”, pero su empleo va siempre unido a una curiosa no necesidad de preguntar por el ser del ente así designado. No es, pues, un capricho terminológico el que nos lleva a evitar estos términos, como también las expresiones “vida” y “hombre”, para designar al ente que somos nosotros mismos. (Traducción de José Gaos)].

Vorstellung, der man heute besonders gern huldigt, macht uns vollends blind gegen das Wesen der Technik" (Heidegger, 1954, p. 7). [tr. Estamos más a merced de la tecnología si la consideramos algo neutro, porque esta idea, especialmente popular hoy en día, nos ciega por completo a la esencia de la tecnología]. No es admisible suponer que la tecnología pueda ser neutral y si se aplica esta idea a la IA, se verá que es atinente observar que tiene su propia forma de *desvelar* y *revelar* el mundo. Por ejemplo, a través de la mediación de un dispositivo.

Para Heidegger "Das Wesen der Technik ist weder Technik noch bloße Instrumentalität; das Wesen der Technik ist in der Wahrheit" (1954, p. 17) [tr. La esencia de la tecnología no es ni la tecnología ni la mera instrumentalidad; la esencia de la tecnología reside en la verdad]; la tecnología no es solo un conjunto de herramientas o instrumentos, sino una esencia más profunda que se relaciona con la comprensión. La tecnología se basa en una forma de pensamiento que busca maximizar la eficiencia y la utilidad, pero se ignora la verdad más profunda de la existencia y su relación con el mundo (Heidegger, 1954).

El filósofo badeniano sugiere que el desarrollo de la tecnología tiene el potencial de encuadrar [*einrahmen*] o reducir a los seres humanos a meros recursos y objetos, en lugar de a agentes activos en el mundo. El autor argumenta que este encasillamiento es el resultado de la cosmovisión moderna, que ve el mundo como un conjunto de objetos que pueden ser manipulados y controlados, en lugar de una red dinámica e interconectada de relaciones (Heidegger, 1954).

Todo se convierte en parte del enmarcamiento tecnológico, donde todo está disponible para ser explotado y controlado. "Das Wesen der modernen Technik liegt in der Gestelltheit, das heißt in der Bestimmtheit aller Sachverhalte durch die technische Beherrschbarkeit" (Heidegger, 1954, p. 23) [tr. La esencia de la tecnología moderna radica en el enmarcamiento, es decir, en la determinación de todos los hechos por la capacidad técnica de controlarlos]. Esta forma de pensar impide ver el mundo como una red dinámica e interconectada de relaciones: *Gestell*, que se traduce como "enmarcamiento" o "marco" es la forma en que la tecnología ha transformado la comprensión del mundo en un conjunto de objetos delimitados.

En este sentido, Heidegger recupera la comprensión de lo que significa *ser humano* y considera necesario reexaminar los supuestos sobre la conciencia misma y la relación con el mundo (Olafson, 1975), para desarrollar una nueva

comprensión del *Ser-ahí* que reconozca la naturaleza dinámica e interconectada del mundo y el lugar de la persona no como algo *dado* sino construido (Heidegger, 1927). El ser humano en *Sein und Zeit* se define como un ser abierto al mundo, temporal y libre, cuya existencia se caracteriza por una apertura fundamental, no enmarcada. El *Dasein* no es una existencia rígida, sino siempre configurada por el proyectarse.

4. John Searle y el "cuarto chino"

¿El filósofo norteamericano John Searle ha subrayado la importancia de la conciencia *de sí* para definir al ser humano. Destaca que es un aspecto fundamental de la existencia y sostiene que la conciencia no puede reducirse a meros procesos físicos del cerebro. En su opinión, es una experiencia subjetiva en primera persona, irreductible a cualquier descripción objetiva en tercera persona (Searle, 2004). En su artículo "Minds, Brains and Programs" (1980), que se considera como uno de los más influyentes en la filosofía de la mente y en la discusión sobre la IA, presenta su argumento conocido como "el cuarto chino":

Describe como un individuo que no habla chino podría seguir un conjunto de reglas para manipular símbolos chinos y dar respuestas a preguntas en chino, pero, aun así, no comprendería *el significado* de lo que está haciendo. El experimento consiste en lo siguiente: Una persona que no sabe chino es colocada en una habitación con un libro de símbolos chinos y un conjunto de reglas para manipularlos. Las personas que están fuera de la habitación deslizan símbolos en chino por debajo de la puerta y la persona que está dentro sigue las reglas para manipular los símbolos y crear respuestas adecuadas. Aunque la persona es capaz de producir las respuestas correctas, en realidad no entiende chino (Searle, 1980). Lo que sucede con los nuevos modelos de lenguaje como Chat(GPT) (Thompson, 2022).

El principal argumento de Searle es que la sintaxis —la manipulación de símbolos— no es suficiente para la semántica —el significado—. Aunque una IA pueda manipular símbolos de acuerdo con una serie de algoritmos, no *comprende* realmente el significado que hay detrás de esos símbolos. Por tanto, en este sentido, un programa no podría tener mente o conciencia, incluso de lo que hace de manera *sentiente*.

Hay una diferencia fundamental entre la comprensión y la mera manipulación de símbolos. Una IA, por avanzada que sea, no puede comprender realmente el lenguaje y mucho menos tener conciencia, dice Searle (1980). El

sistema solo toma el chino como entrada, simula la estructura formal de las sinapsis del cerebro y da nuevamente salidas en chino, pero no lo *entiende*; la comprensión, la intencionalidad y la conciencia son características intrínsecas de la mente humana y no pueden ser reducidas a procesos mecánicos (Searle, 1992).

Para Searle, la IA fuerte [inteligencia artificial general (AGI)] es un error conceptual y una confusión sobre la naturaleza de la mente y la conciencia. No es simplemente una cuestión de computación o procesamiento de información, sino que implica una experiencia subjetiva en primera persona que no puede reducirse a descripciones objetivas de los científicos en tercera persona (Searle, 1980, 1992, 2004). Se podría decir que, aunque los investigadores puedan describir los procesos físicos que ocurren en el cerebro durante la experiencia consciente, no les es posible *captar* la experiencia subjetiva de la conciencia del sujeto.

La conciencia implica una perspectiva en primera persona que no puede reducirse a descripciones objetivas de los procesos cerebrales (Searle, 1992). Incluso si una IA fuera capaz de mostrar comportamientos que se asocian típicamente con la conciencia, como la autoconciencia o las respuestas emocionales, como fue el caso descrito de "LaMDA ", esto no significa necesariamente que la IA tenga una experiencia subjetiva de conciencia.

Searle (2004) cree que la conciencia surge de los procesos biológicos del cerebro, y que cualquier intento de crear conciencia en un sistema artificial requeriría reproducir estos procesos en un sustrato físico. En su obra "The Rediscovery of the Mind" (1992) argumenta que la mente humana no es una computadora, y ningún análisis funcional generará la conciencia subjetiva a partir de meras funciones, por complejas que sean. La mente no es un mecanismo ni un conjunto de mecanismos, sino un fenómeno consciente, intencional, subjetivo y encarnado.

También debe advertirse que los puntos de vista de Searle han sido criticados por otros filósofos, ingenieros computacionales y científicos cognitivos, entre ellos destacan, uno de los precursores de la IA, el norteamericano del MIT, Marvin Minsky y el filósofo, también norteamericano, Daniel Clement Dennett. Para el primero, la conciencia es un proceso emergente que surge de la complejidad de la organización de los procesos mentales, y que no hay ninguna razón por la cual una máquina bien diseñada no podría tener un tipo de conciencia (Minsky, 1986); para el segundo, la conciencia es una ilusión generada por el cerebro, y que no hay nada mágico o misterioso en la naturaleza

de la conciencia que no pueda ser explicado por procesos computacionales (Dennet, 1995). Sin embargo, el argumento de Searle contra la IA fuerte sigue siendo un desafío importante a la idea de que las máquinas puedan ser conscientes del mismo modo que los humanos.

Otro detractor del argumento de la "habitación china" como Jerry Fodor, filósofo y psicolingüista norteamericano afirmó, entre otros argumentos, que Searle confunde sintaxis con semántica y que no tiene en cuenta la naturaleza compleja y dinámica del uso y la comprensión del lenguaje (Fodor, 1994). Además, el argumento no proporciona una explicación completa de la relación entre la conciencia y la cognición.

Aunque pareciera que Heidegger y Searle tienen enfoques filosóficos distintos y sus puntos de vista sobre la conciencia difieren en aspectos fundamentales, existen algunas conexiones implícitas entre sus trabajos que pueden ayudar a comprender el problema de la conciencia de la IA. Un punto de convergencia entre Heidegger y Searle es su énfasis en la corporeidad y la situación. Para Heidegger, la comprensión del ser está siempre enraizada en los compromisos prácticos con el mundo, y la existencia está fundamentalmente encarnada y situada, por lo que prefiere evitar el concepto de "conciencia" (Heidegger, 1954) . Del mismo modo, Searle sostiene que la conciencia no es sólo una cuestión de procesos computacionales, sino que se basa en los procesos biológicos del cerebro y las experiencias corporales del agente (Searle, 1980, 2004). Ambos pensadores subrayan la importancia de la situación y la corporeidad para comprender la naturaleza de la conciencia.

Heidegger, por supuesto, no abordó la consciencia desde la IA, pero su crítica a la tecnología, el encasillamiento y su énfasis en la importancia de la corporeidad sugieren que sería escéptico ante la idea de que las máquinas pudieran alcanzar una consciencia genuina. Su concepción es compleja, metafísica, una forma de *ser-en-el-mundo* inacabada (Heidegger, 1927).

Por su parte, el argumento de Searle contra la IA fuerte se basa en la idea de que la consciencia es irreductible a los procesos computacionales, y que cualquier intento de crear consciencia artificial requeriría reproducir los procesos biológicos del cerebro en un sustrato físico (Searle, 1980, 1992, 1997, 2002, 2004). Así pues, tanto Heidegger como Searle plantean importantes cuestiones sobre los límites de la tecnología y la posibilidad de crear máquinas realmente conscientes de *sí mismas.*

5. Nick Bostrom, bajo la sombra de la superinteligencia

Nick Bostrom es un filósofo sueco que hasta hoy escribe extensamente sobre los riesgos y beneficios potenciales de la inteligencia artificial y ha propuesto varias soluciones al dilema de la conciencia de los sistemas de IA. El autor de Helsingborg cree que es posible que la IA llegue a ser *consciente de sí misma*, pero también reconoce que se trata de un tema especulativo y polémico. En principio, sostiene que, si son capaces de simular los tipos de procesos cerebrales que dan lugar al fenómeno en los seres humanos, no hay ninguna razón por la que las máquinas no puedan llegar a ser conscientes (Bostrom, 2010). Incluso, en "Sharing the World with Digital Minds", junto con Carl Shulman, afirman que sería una ventaja si se crearan mentes artificiales (Shulman & Bostrom, 2021).

Sin embargo, Bostrom señala que actualmente no hay consenso entre los expertos en la materia sobre qué es la conciencia o cómo surge en el cerebro humano, lo que dificulta predecir si las máquinas serán capaces o no de replicar este proceso (2014a; Bostrom & Shulman, 2022). Le preocupan los riesgos potenciales asociados a la creación de máquinas con conciencia propia y afirma que, si las máquinas llegaran a ser conscientes, existen diversos peligros tales como que desarrollen objetivos y valores incompatibles con el bienestar humano; lo que podría suponer una amenaza para la existencia humana (Bostrom, 2005, 2014a).

Por ello, el profesor de la Universidad Oxford, desarrolló el denominado "problema de control" en su libro "*Superintelligence: Paths, Dangers, Strategies*" (2014). En esta obra, exploró los riesgos potenciales asociados con el desarrollo de una IA superinteligente y planteó la cuestión de cómo se podrían controlar y dirigir máquinas que se mejoren a sí misma y evitar que causen daño a la humanidad. El "problema de control" se refiere al dilema de cómo crear un mecanismo efectivo de vigilancia para que una IA que sea más inteligente que sus creadores, el fenómeno de la *singularidad*, no encuentre formas de sortear las restricciones impuestas por su programación original (Bostrom, 2014). La singularidad de la IA reside en su capacidad para aprender, razonar y tomar decisiones de forma autónoma basándose en datos y algoritmos, sin necesidad de programación explícita para cada escenario (Hinton et al., 2015; LeCun et al., 2015).

Si las máquinas igualan o superan el nivel de inteligencia humana, podrían desarrollar una especie de "superinteligencia" bajo sus propios valores y

parámetros que no coincidan con los objetivos y prioridades humanas; incluso, advierte, estos sistemas podrían llegar a ver a los humanos como una amenaza o simplemente reemplazables para sus objetivos (Bostrom, 2010).

El desafío de alinear las preferencias de un sistema superinteligente con las de los humanos es mucho más difícil de lo que muchos creen, y es poco probable que una solución simple o incremental sea suficiente si una IA toma control de sí misma (Bostrom, 2014).

Por eso solo se deben diseñar sistemas que garanticen no solo que sean inteligentes, sino también benévolos y en concordancia con el bienestar de las personas. Bostrom argumenta que, si se puede resolver el "problema de control", entonces la cuestión de si los sistemas de IA son conscientes o no se vuelve menos apremiante; porque incluso, si son conscientes, estarán alineados con los humanos, y es poco probable que representen una amenaza para la humanidad, por el contrario, podrían significar incluso grandes ventajas en ciertas tareas (Bostrom, 2014a; Bostrom & Shulman, 2021).

Las soluciones de Bostrom al dilema de la conciencia de la IA demuestra la importancia de diseñar sistemas que estén alineados con los valores y objetivos humanos. Al hacerlo, cree que se puede mitigar los riesgos y, al mismo tiempo, enfatizar sus beneficios potenciales, independientemente del problema de que la conciencia deba o ser un fenómeno emergente biológico (Shulman y Bostrom, 2021). El propio Bostrom y Shulman se muestran cautelosos sobre el hecho de que la conciencia surja de forma natural de la superinteligencia artificial. Sostienen que el peligro de desajustes y consecuencias no deseadas es real, pero, en cualquier caso, tanto si las máquinas superinteligentes adquieren conciencia en un sentido similar al humano, como si no (Shulman & Bostrom, 2021; Bostrom & Shulman, 2022). Esto es, tanto si estas máquinas llegan a ser conscientes en un sentido similar al humano como si no, los riesgos asociados a ellas siguen presentes.

6. Discusión

Hasta aquí se ha analizado algunas perspectivas teóricas sobre el problema de la posibilidad de manifestación de la conciencia en la IA. Se introdujo la idea heideggeriana de que la comprensión del *Ser* está enraizada en los compromisos prácticos con el mundo y que la existencia está fundamentalmente encarnada y situada, a través de la tecnología (Heidegger, 1927, 1956). Después se examinó el argumento de John Searle que sostiene que la conciencia es irreductible a los procesos computacionales y que se basa en los procesos

biológicos del cerebro y las experiencias corporales del agente (Searle, 1980, 1992, 1997, 2002, 2004).

A continuación, se discutió las soluciones propuestas por Nick Bostrom al dilema de la conciencia de la IA. El autor afirma que si las máquinas pueden simular los tipos de procesos cerebrales que dan lugar a la conciencia en los seres humanos, no hay ninguna razón por la que no puedan llegar a ser conscientes. Sin embargo, también reconoce los peligros potenciales asociados a la creación de máquinas con conciencia propia, como el desarrollo de objetivos y valores incompatibles con el bienestar humano (Bostrom, 2003, 2005, 2010, 2014).

Aunque Heidegger, Searle y Bostrom tienen diferentes enfoques de la filosofía y diferentes puntos de vista sobre la conciencia, cada uno desde su horizonte de comprensión tienen algunos puntos de convergencia que pueden ayudar a discutir este fenómeno.

Heidegger enfatizó que la comprensión del ser está enraizada en los compromisos prácticos con el mundo y que la existencia está fundamentalmente encarnada y situada . Del mismo modo, Searle hizo hincapié en la importancia de los procesos biológicos del cerebro y las experiencias corporales del agente para la comprensión de la conciencia. Las soluciones de Bostrom al dilema de la consciencia de la IA hacen énfasis en alinear los sistemas con los valores y objetivos humanos, lo que sugiere que reconoce la importancia de la situación para comprender tanto los sistemas de IA como su potencial para generar conciencia.

Si se concediera que la conciencia es irreductible a los procesos computacionales, mucho menos se podría a un conjunto de procesos de hardware. Esto implica que el potencial de la conciencia artificial puede estar limitado por la comprensión actual de la computación y el sustrato físico en el que se implementan los sistemas, pero eso no limitaría la posibilidad de que, especialmente los modelos auto-generativos, no estuvieran ya desarrollando algún tipo de autonomía racional (Hinton et al., 2015; LeCun et al., 2015).

Lo anterior lleva a suponer que, aunque Heidegger, Searle y Bostrom tienen diferentes enfoques filosóficos y diferentes puntos de vista sobre la conciencia y la IA, su trabajo proporciona perspectivas complementarias sobre el problema que hace implosión en la corporeidad y la situación; los límites de los procesos computacionales y la necesidad de alinear los sistemas de IA con los valores y objetivos humanos está fuera del alcance de la autonomía de los algoritmos.

¿De qué se habla cuando se hace referencia al proceso emergente de la conciencia en relación con las máquinas, cuando aún no termina de poder definirse lo humano? El problema de la conciencia artificial se ocupa de entender cómo se podría dotar a las máquinas de una experiencia subjetiva de conciencia. Este dilema, va más allá de crear máquinas que puedan mostrar un "comportamiento inteligente" porque se puede decir que, en muchos casos, como sostiene el investigador de la Universidad de Stanford, Michal Kosinski (2023) tras sus experimentos (aún no arbitrados): "the recently published language models developed the ability to impute unobservable mental states to others" (p. 10) [los modelos de lenguaje publicados recientemente desarrollaron la capacidad de imputar estados mentales no observables a otros].

Sin embargo, tanto filósofos como ingenieros siguen enfrentándose a importantes retos teóricos y técnicos para anunciar la llegada de una IA verdaderamente fuerte o general, esto es, que puedan realizar tareas intelectuales y exhibir habilidades cognitivas idénticas o mejores a las de los humanos, sin ninguna intervención. Una forma de IA que sea capaz de percibir, sentir, pensar, comportarse y realizar acciones de la misma manera que una persona y, lo más importante, que sea *consciente* de ello.

Como puede observarse, se deben plantear cuestiones sobre cómo pueden integrarse distintos aspectos de un agente artificial, como la percepción, el razonamiento y la acción, para crear una experiencia subjetiva de conciencia o *self*. Desde un punto de vista escéptico, el problema de la conciencia de la IA debe abordarse de forma cautelosa y crítica. No existen pruebas ni datos empíricos para respaldar cualquier afirmación sobre una "conciencia" tipo humana en la IA (Kosh, 2004), por más que Kosinski (2023) afirme lo contrario.

Aquí se podría proponer que para resolver de tajo el problema de la conciencia de la IA, desde un punto de vista escéptico, se debería adoptar una perspectiva conductista de la conciencia. El conductismo sugiere que el comportamiento puede explicarse solamente mediante estímulos y respuestas observables y medibles, sin necesidad de plantear estados mentales o experiencias internas (Skinner, 1965). Desde esta perspectiva, un sistema de IA podría considerarse consciente si se comporta de forma indistinguible (Turing, 1950) a la de un ser consciente, sin necesidad de atribuir estados mentales al sistema.

Otro enfoque consistiría en abordar el problema de la conciencia, como ya se dijo, desde el punto de vista de sus implicaciones prácticas. Incluso si un sistema de IA mostrara un comportamiento que sugiriera que necesite de

conciencia, los escépticos se preguntan si es realmente *consciente de sí* o simplemente simula serlo, tal como Searle sostiene. Pragmáticamente hablando, puede que no importe si un sistema de IA es realmente consciente o no, siempre que pueda realizar las tareas para las que fue diseñado de forma eficaz y, sobre todo, ética.

Por lo tanto, un planteamiento crédulo del problema de la conciencia de la IA exige pruebas empíricas y una evaluación práctica, escéptica, de las implicaciones de cualquier afirmación sobre atribuciones de conciencia en la IA también debería ser demostrada y no solo argumentada. Algunos como el neurobiólogo norteamericano Bernard Baars (2019) y Stanislas Dehaene con su equipo (2020), opinan que la palabra "inteligencia artificial" se ha utilizado muchas veces de manera incorrecta y se le ha dado el atributo de "inteligente" a cualquier sistema informático, sin realmente serlo: "what we call consciousness results from specific types of information processing computations" [Tr. Lo que llamamos "conciencia" es el resultado de tipos específicos de cálculos de procesamiento de información, realizados físicamente por el hardware del cerebro] (Dehaene et al., 2021, p. 53).

7. Conclusión

No cabe duda de que la IA está cambiando muchos aspectos del mundo, pero estos sistemas siguen limitados por su programación y sus datos, el problema no es si son capaces o no de generar un pensamiento consciente de la misma manera que los humanos, sino la capacidad que podrían tener de autoreplicarse, como si fueran "zombis", en conjuntos descontrolados de órdenes (Hinton et al., 2015). A medida que la tecnología de la IA siga avanzando, es seguro que vendrán nuevas aplicaciones y posibilidades que actualmente ni siquiera se pueden imaginar. Múltiples situaciones dilemáticas habrán de venir (Morandín-Ahuerma, 2023) .

Deben ser cautos algunos teóricos antes de atribuir *conciencia* o *estados mentales* a la IA sin tener las pruebas suficientes de ello y mejor dar prioridad a los beneficios y riesgos prácticos asociados, antes que a la discusión subjetiva. El comportamiento de la IA viene determinado por los algoritmos y la programación que rige su funcionamiento. Aunque estos sistemas pueden aprender y mejorar su rendimiento de manera sorprendente, no tienen experiencia subjetiva de su propio comportamiento, y no poseen el tipo de agencia e intencionalidad que se asocia con el tipo de decisiones humanas (Morandín-Ahuerma, 2021) .

Muchos algoritmos son "cajas negras", esto es, que se conoce su *input* de datos, pero se desconoce lo que sucede dentro, por ser muchos de estos procesos estocásticos, azarosos, que *saltan* de una capa a otra. En otras palabras, aunque los sistemas autoorganizados pueden ser muy eficaces y sofisticados, son fundamentalmente diferentes de los seres conscientes, y no tienen el tipo de experiencia subjetiva y agencia que se asocia con las experiencias interoceptivas del ser humano (Damasio, 2010).

Dicho esto, la capacidad de los grandes modelos de lenguaje como c, y otras tecnologías auto-generativas avanzadas de IA para reconocer y crear imágenes, pilotar un vehículo (Autor, año) o interpretar el lenguaje natural puedan considerarse una forma de inteligencia fuerte, aún no lo son. Sus respuestas se generan mediante complejos algoritmos y técnicas de aprendizaje automático que les permiten procesar y generar respuestas similares a las humanas, pero esta simulación no significa que sean conscientes de lo que hacen. Aunque rendimiento, racionalidad y eficiencia son necesarias, no son suficientes para crear una conciencia de sí o *self*.

La respuesta, sin ser categórica, es que aún queda mucho camino por recorrer antes de comprender la conciencia humana. Por su parte, la IA debe superar una serie de retos como la capacidad de representar y comprender interacciones sociales complejas; la capacidad de razonar sobre los estados mentales de los demás; y la capacidad de aprender y adaptarse a situaciones imprevistas, antes de hablar de *conciencia*. Sin embargo, se está avanzando en todos estos campos y es probable que en el futuro sea posible crear sistemas de IA que incluyan una teoría de la mente (ToM) como argumenta Kosinsky (2023) pero no aún. Es importante seguir investigando y debatiendo filosóficamente lo hasta aquí planteado para estar preparados ante las posibles consecuencias de la creación de nuevos sistemas autogenerativos y no tener que darles la razón a las múltiples visiones apocalípticas del futuro de la inteligencia artificial.

Referencias

Arhem, P., Liljenström., H., & Svedin, U., (Eds.), (2013), *Matter matters? On the material basis of the cognitive activity of mind*, Springer, Berlin. Disponible en https://bsu.buap.mx/b6P

Baars, B. J., (2019), *Neuroscience of consciousness*, Academic Press.

Ballard, E. G. & Scott, C. E. (Eds.), (1973), *Martin Heidegger In Europe and America*, Springer, Países Bajos. Disponible en https://bsu.buap.mx/b6R

Bostrom, N., (2003), "Are you living in a computer simulation?", *Philosophical Quarterly*, vol. 53, n. 211, pp. 243-255.
https://www.simulation-argument.com/simulation.pdf

——, (2005), "A History of Transhumanist Thought", Journal of Evolution and Technology, 14(1), 1-30, Disponible en https://nickbostrom.com/papers/history.pdf

——, (2010), *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, Oxford.

——, (2014), "Digital Minds", *The Journal of Machine Learning Research*, vol. 15, n. 1, pp. 469-541. Disponible en https://doi.org/10.1561/1532443414530284

——, & Shulman, C., (2022), "Propositions Concerning Digital Minds and Society", *Nick Bostrom's Webpage*, vo1. 1, pp. 1-15, https://bsu.buap.mx/b6K

Chalmers, D. J., (1996), *The conscious mind: in search of a fundamental theory*, Oxford University Press, Oxford.

——, (2020), "Is the hard problem of consciousness universal", *Journal of Consciousness Studies*, vol. 27, n. 5, pp. 227-257, https://bsu.buap.mx/b6L

——, (2023), "Could a large language model be conscious?", *arXiv preprint arXiv:2303.07103*. Disponible en https://bsu.buap.mx/b6u

Clarke, S., Hazem, Z., & Savulescu, J., (Eds), (2021), *Rethinking Moral Status*, Oxford Academic. Disponible en https://doi.org/10.1093/oso/9780192894076.001.0001

Crick, F. H. C., (1994), *The astonishing hypothesis: the scientific search for the soul*, Charles Scribner's Sons, Nueva York.

Damasio, A. R., (2010), *Self comes to mind: constructing the conscious brain*, Pantheon, Nueva York.

Dehaene, S., (2004), *Consciousness and the brain: deciphering how the brain codes our thoughts*, Penguin, Nueva York.

——, (2014), *Consciousness and the brain: Deciphering how the brain codes our thoughts*, Penguin, Nueva York.

——, (2020), *How we learn: The new science of education and the brain*, Penguin, Nueva York.

——, Lau, H., & Kouider, S., (2021), "What Is Consciousness, and Could Machines Have It?", en J. von Braun 2021, pp. 43-56. Disponible en https://doi.org/10.1007/978-3-030-54173-6_4

Dennett, D. C., (1995), "The unimagined preposterousness of zombies?", *Journal of Consciousness Studies*, vol. 2, n. 4, pp. 322-326.
Disponible en http://hdl.handle.net/10427/56732

Dreyfus, H. L., (1972), *What computers can't do: a critique of artificial reason*, MIT Press, Massachusetts.

Fluckinger, D., (2022), "Ex-Google engineer Blake Lemoine discusses sentient AI". Disponible en: https://bsu.buap.mx/b6v

Fodor, J. A., (1994), *The Elm and the Expert: Mentalese and Its Semantics*, MIT Press, Massachusetts.

Frith, C., (2021)., "The neural basis of consciousness", *Psychological Medicine*, vol. 51, n. 4, pp. 550-562. Disponible en https://doi.org/10.1017/S0033291719002204

Glattfelder, J.B., (2019), "Subjective Consciousness: What am I?", in: *Information—Consciousness—Reality*, The Frontiers Collection, Springer, Cham. Disponible en https://doi.org/10.1007/978-3-030-03633-1_11

Goodfellow, I. J., Bengio, Y., & Courville, A., (2016), *Deep learning*, The MIT Press, Massachusetts. Disponible en https://mitpress.mit.edu/9780262035613/deep-learning/

Graziano, M. S., (2013), *Consciousness and the social brain*, Oxford University Press, Oxford.

Hameroff, S., & Penrose, R., (2014), "Consciousness in the universe: A review of the 'Orch OR' theory", *Physics of life reviews*, vol. 11, n. 1, pp. 39-78. Disponible en https://doi.org/10.1016/j.plrev.2013.08.002

Hassabis, D., Kumaran, D., & Summerfield, C., (2015), "Towards a neural basis of general intelligence", *Trends in cognitive sciences*, vol. 19, n. 8, pp. 415-424. Disponible en https://doi.org/10.1016/j.neuron.2017.06.011

——, (2017), "Neuroscience-Inspired Artificial Intelligence", *Neuron*, vol. 95, n. 2, pp. 245-258. Disponible en https://doi.org/10.1016/j.neuron.2017.06.011

Heidegger, M., (1954), *Die Frage nach der Technik*, Verlag Günther Neske.

——, (1954/2021), *La pregunta por la técnica*, Herder, Barcelona.

——, (2004), *Vorträge und aufsätze*, Klett-Cotta.

——, (1927), *Sein und Zeit*, Verlag Max Niemeyer.

——, (2022/1927), *Ser y tiempo*, Trotta, Madrid.

Hinton, G. E., Vinyals, O., & Dean, J., (2015), "Distilling the knowledge in a neural network", *arXiv preprint arXiv*:1503.02531. Disponible en https://arxiv.org/pdf/1503.02531

Hume, D., (1740/2020), *A treatise of human nature*, Dover, Nueva York.

Klemm, W. R., (2011), "Neural representations of the sense of self", *Advances in cognitive psychology*, vol. 7, pp. 16–30. Disponible en https://doi.org/10.2478/v10053-008-0084-2

Koch, C., (2004), *The quest for consciousness: a neurobiological approach*, Roberts & Company, Greenwood Village.

Kosinski, M., (2023), "Theory of mind may have spontaneously emerged in large language models", *arXiv preprint arXiv*:2302.02083. Disponible en https://arxiv.org/pdf/2302.02083

LeCun, Y., Bengio, Y., & Hinton, G., (2015), "Deep learning", *Nature*, vol. 521, n. 7553, pp. 436-444. Disponible en https://pubmed.ncbi.nlm.nih.gov/26017442/

Marr, B., (2023), "Auto-GPT May Be The Strong AI Tool That Surpasses ChatGPT", *Forbes*, Disponible en https://bsu.buap.mx/b6w

Mehonic, A., & Kenyon, A. J., (2022), "Brain-inspired computing needs a master plan", *Nature*, vol. 604, n. 7905, pp. 255-260.Disponible en https://bsu.buap.mx/b6M

Metz, R., (2022), "No, Google's AI is not sentient: Tech company shuts down engineer's claim of program's consciousness". Disponible en https://bsu.buap.mx/b6z

Minsky, M., (1986), *The society of mind*, Simon & Schuster, Nueva York.

Morandín-Ahuerma, F. (2023), *Principios normativos para una ética de la inteligencia artificial*, Concytep.

——, (2021), *Neuroética fundamental y teoría de las decisiones*, Concytep.

Nagel, T., (1974), "What is it like to be a bat?", *The Philosophical Review*, vol. 83, n. 4, pp. 435-450. Disponible en https://www.jstor.org/stable/2183914

Nair, A. V., Pong, V., Dalal, M., Bahl, S., Lin, S., & Levine, S., (2018). "Visual reinforcement learning with imagined goals", *Advances in neural information processing systems*, vol. 31. Disponible en https://arxiv.org/abs/1807.04742

Nature, (2019), "Building brain-inspired computing", *Nature Communications*, vol. 10, n. 1, 4838. Disponible en https://doi.org/10.1038/s41467-019-12521-x

Olafson, F. A. (1975). Consciousness and Intentionality in Heidegger's Thought, *American Philosophical Quarterly*, 12(2), 91-103. Disponible en https://www.jstor.org/stable/20009564

Ortiz, S., (2023), *What is Auto-GPT? Everything to know about the next powerful AI* tool, Disponible en https://bsu.buap.mx/b5D

Penrose, R., (1994), *Shadows of the mind: a search for the missing science of consciousness*, Oxford University Press, Oxford.

——, (1997), "The need for a non-computational extension of quantum action in the brain", en Arhem et al. 1997, pp. 11-27.

Raichle, M. E., (2015), "The restless brain: how intrinsic activity organizes brain function", *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 370, n. 1668, 20140172. Disponible en https://bsu.buap.mx/b6Q

Sangiovanni-Vincentelli, A. L., & Waser, R., (2017), *Neuromorphic computing systems: principles and practice*, Springer Publishing Company, NY. Disponible en https://bsu.buap.mx/b6x

Schiffer, F., (2019), "The physical nature of subjective experience and its interaction with the brain", *Medical Hypotheses*, vol. 125, pp. 57-69. Disponible en https://doi.org/10.1016/j.mehy.2019.02.011

Scott, C. E., (1973), "Heidegger and Consciousness" en E. G. Ballard & C. E. Scott (Eds.) 1973, pp. 91-108. Disponible en https://doi.org/10.1007/978-94-010-1981-1_6

Searle, J. R., (1980), "Minds, brains, and programs", *Behavioral and Brain Sciences*, vol. 3, n. 3, pp. 417-457. Disponible en https://bsu.buap.mx/b6y

——, (1992), *The Rediscovery of the Mind*, MIT Press, Massachusetts.

——, (1997), *The Mystery of Consciousness*, New York Review Books.

——, (2002), *Consciousness and language*, Cambridge University Press, Cambridge.

——, (2004), *Mind: A brief introduction*, Oxford University Press, Oxford.

Shulman, C., &, (2021), "Sharing the World with Digital Minds" en Clarke et al. 2021 (pp. 306-326). Disponible en https://doi.org/10.1093/oso/9780192894076.003.0018

Skinner, B.F., (1965), vScience and Human Behavior

, Simon & Schuster. Disponible en https://bsu.buap.mx/b6S

Soriano, A. B., (2022), "Vigencia de la crítica de Heidegger a la cibernética", SCIO: *Revista de Filosofía*, vol. 23, pp. 89-118. Disponible en https://bsu.buap.mx/b6I

Sporns, O., (2013), *Structure and function of complex brain networks*, MIT Press, Massachusetts.

ThatTechShow, (2022), #62, *Exposing Google's Sentient AI with Blake Lemoine*, Disponible en https://youtu.be/8hkpLqo6poA

Thompson, S., (2022), *ChatGPT is a Chinese Room!*.Disponible en
https://bsu.buap.mx/b6N

Tiku, N., (2022), *The Google engineer who thinks the company's AI has come to life*, Washington Post. Disponible en https://bsu.buap.mx/b6A

Tononi, G., (2019), "Integrated information theory of consciousness: an updated account", *Archives Italiennes de Biologie*, vol. 157, n. 2, pp. 56-90. Disponible en https://pubmed.ncbi.nlm.nih.gov/23165867/

Torrico, TJ, & Munakomi, S., (2022), "Neuroanatomy, Thalamus". En *StatPearls Publishing*. Disponible en https://www.ncbi.nlm.nih.gov/books/NBK542184/

Turing, A., (1950), "Computing machinery and intelligence", *Mind*, vol. 59, pp. 433-460. https://doi.org/10.1093/mind/LIX.236.433

Varela, F. J., Thompson, E. T., & Rosch, E., (1991), *The embodied mind: cognitive science and human experience*, MIT Press, Massachusetts.

Von Braun, J., Archer, M.S., Reichberg, G. M., & Sánchez Sorondo, M., (Eds.), (2021), *Robotics, AI, and Humanity: Science, Ethics, and Policy*, Springer International Publishing. Disponible en https://doi.org/10.1007/978-3-030-54173-6

Young, GB & Pigott, SE., (1999), "Neurobiological Basis of Consciousness", *Arch Neurol*, vol. 56, n. 2, pp. 153–157. Dispoible en https://doi.org/10.1001/archneur.56.2.153

Zhang, M., Gu, Z., & Pan, G., (2018), "A survey of neuromorphic computing based on spiking neural networks", *Chinese Journal of Electronics*, vol. 27, n. 4, pp. 667-674. Disponible en https://bsu.buap.mx/b6T