

Más allá del algoritmo: conciencia, subjetividad y límites de la emulación artificial

Beyond the algorithm: consciousness, subjectivity, and limits of artificial emulation

Arturo Salazar Palomares 

Universidad Anáhuac Querétaro, México

arturo.salazar80@anahuac.mx

Recepción: 21 de agosto de 2025

Aceptación: 9 de octubre de 2025

Resumen

En este artículo se examinan los límites filosóficos y fenomenológicos de la idea de transferir la conciencia humana a una máquina. Existen teorías como el Espacio de Trabajo Neuronal Global (ETNG), la Teoría de la Información Integrada (TII) y la Emulación Cerebral Completa (ECC) que proponen modelos funcionales de la conciencia; sin embargo, estas aproximaciones no logran captar la dimensión subjetiva, encarnada y afectiva de la misma. A través de un diálogo entre la neurociencia cognitiva, la filosofía de la mente y la fenomenología se sostiene que la conciencia es más que solo información procesada: es una experiencia vivida, situada y corpórea. Por ello, se afirma que la transferencia completa de la conciencia a una máquina no es inviable por cuestiones técnicas, sino por una comprensión reduccionista de la condición humana. Se concluye que la conciencia no puede disociarse ni del cuerpo ni de la historicidad vivida, sin despojarla de su esencia.

Palabras clave: conciencia, fenomenología, identidad personal, reduccionismo, transferencia mental

Abstract

In this article, the philosophical and phenomenological limits of the idea of transferring human consciousness to a machine are examined. There are theories such as the Global Neuronal Workspace (GNW), the Integrated Information Theory (IIT), and the Whole Brain Emulation (WBE) that propose functional models of consciousness, however, it is argued that these approaches fail to capture its subjective, embodied, and affective dimensions. Through a dialogue between cognitive neuroscience,

philosophy of mind, and phenomenology, it is maintained that consciousness is more than processed information: it is a lived, situated, and embodied experience. Therefore, it is argued that the complete transfer of consciousness into a machine is not unfeasible due to technical issues, but rather due to a reductionist understanding of the human condition. It is concluded that consciousness cannot be dissociated from the body or from lived historicity, without stripping it of its essence.

Keywords: *consciousness, mental transfer, personal identity, Phenomenology, reductionism*

De la ciencia ficción a la realidad, el dilema de Roy Batty

La idea de que la conciencia humana puede ser transferida a una máquina ha dejado de ser una simple especulación futurista para convertirse en un punto de fricción entre la neurociencia, la inteligencia artificial y la filosofía de la mente. Lo que en el pasado parecía un argumento exclusivo de la ciencia ficción, hoy se discute seriamente en ámbitos académicos y tecnológicos, impulsado por teorías científicas de la conciencia. Algunas de ellas presentan un enfoque neurocognitivo (Dehaene, 2014), otras buscan explicarse a través de un proceso matemático-fenomenológico (Tononi, 2008), o mediante un enfoque tecnológico-filosófico dentro de la corriente del Transhumanismo (Kurzweil, 2024). Desde una perspectiva esencialmente funcionalista, estas propuestas sostienen que la conciencia podría entenderse como un proceso de integración y distribución de información en redes neuronales, la cual es susceptible de ser replicada en soportes artificiales. Sin embargo, estas hipótesis enfrentan diversas objeciones filosóficas, particularmente en relación con la experiencia subjetiva y, en especial, con los *qualia*, entendidos como los aspectos fenomenológicos inefables de las sensaciones y percepciones. Autores como Chalmers (1996), Nagel (1974), Block (2007) y más recientemente Gazzaniga (2018), han subrayado los límites explicativos de dichos modelos para tratar de dar cuenta de la conciencia en su totalidad.

A partir de este contexto, se plantea la pregunta central de este artículo sobre la posibilidad de transferir plenamente la conciencia humana a una máquina. La hipótesis que aquí se sostiene es que, si bien ciertos procesos cognitivos como la atención, la memoria o la toma de decisiones pueden ser simulados computacionalmente, existen límites filosóficos y fenomenológicos insalvables que impiden una transferencia completa de la misma. Entre los más importantes destacan la dimensión subjetiva de la experiencia humana, la constitución corpórea de la mente y la construcción narrativa de la identidad personal. Para tratar de descifrar si es posible esta transferencia, se propone realizar un análisis teórico y crítico

que contraste las propuestas científicas actuales con los enfoques filosóficos de la conciencia fenoménica y del cuerpo vivido. Este enfoque dialógico e interdisciplinario tiene como objetivo confrontar los argumentos de la neurociencia cognitiva con los de la fenomenología y la filosofía de la mente para ofrecer una evaluación teórica sobre las posibilidades reales y los límites del proyecto transhumanista de digitalización del yo.

Antes de seguir avanzando, es importante puntualizar que se usará la palabra *conciencia* para referirse a la capacidad del ser humano de *darse cuenta* o *reconocer algo*, ya sea del mundo exterior, como objetos, situaciones o cualidades; o bien, de su mundo interior, como pensamientos, emociones o cambios en su propio yo. Esta definición abarca el concepto que comúnmente entendemos como *estar consciente* de algo, y se alinea con la formulación ofrecida por José Ferrater Mora en su *Diccionario de Filosofía* (1965). Ferrater señala que, a lo largo de la historia del pensamiento, el concepto de conciencia ha sido interpretado de formas muy diversas, desde la *autoconciencia reflexiva* en la filosofía moderna, hasta la *conciencia moral* como voz interior que distingue el bien del mal, y más recientemente, como la *experiencia subjetiva* que acompaña nuestros estados mentales. Comprender estas distintas aproximaciones permite situar mejor los debates actuales sobre la conciencia en campos como la neurociencia, la inteligencia artificial y la filosofía de la mente.

Uno de los ejes más relevantes del pensamiento filosófico ha sido la distinción entre una *concepción cosista* de la conciencia, que la entiende como sustancia o facultad autosuficiente, y una *concepción intencional*, que la define como *función dirigida* a algo distinto de sí misma. Mientras que en ciertos enfoques antiguos y medievales predominó la idea de la conciencia como un atributo sustancial del alma, con la modernidad emergen definiciones centradas en la *auto-transparencia* del yo, es decir, que el sujeto es consciente de sí mismo de forma directa, *sabe que sabe* o se percibe a sí mismo pensando, sintiendo o actuando.

De acuerdo con Ferrater (1965), en el siglo xx, la fenomenología consolidó la comprensión de la conciencia como intencionalidad. “Brentano concibe la conciencia como *intencionalidad* y, por lo tanto, hace de la conciencia algo que no es continente ni contenido, sino mera proyección y referencia a aquello que es mentado” (p. 324). De toda esta diversidad de autores y líneas de pensamiento se desprende un rasgo común que será fundamental en el curso de la presente investigación: “la conciencia constituye el principio de unidad de la experiencia” (p. 326). Aquello que permite articular percepciones, recuerdos y expectativas en un horizonte coherente.

Aquí se asumirá precisamente esta significación intencional-fenomenológica de la conciencia. Más que una sustancia aislada o una mera función cognitiva, la conciencia será entendida como la capacidad del sujeto de estar *dirigido hacia el mundo* y de unificar sus experiencias en una *vivencia coherente*. Esta definición resulta la más pertinente para examinar los límites de la emulación artificial pues mientras los sistemas computacionales pueden procesar información, lo que está en cuestión es si pueden replicar esa intencionalidad y unidad vivida que constituyen la base de la subjetividad.

Asimismo, en el marco de este trabajo, también será importante establecer que el término *máquina* no se limita a un artefacto mecánico o a un simple sistema de cómputo, sino que alude a un dispositivo capaz de implementar en un *soporte no biológico* arquitecturas funcionales semejantes a las del cerebro humano. Proyectos como el *Whole Brain Emulation* (WBE), documentado por Kurzweil (2012), apuntan precisamente a la posibilidad de mapear con precisión la conectividad cerebral y reproducir su dinámica causal en otro medio. Dehaene (2014), dentro del marco del *Global Neuronal Workspace* (GNW), sostiene que la conciencia emerge de interacciones distribuidas en redes neuronales de largo alcance y, en principio, una simulación completa de estas redes podría ejecutarse en una máquina digital.

En esta misma línea de pensamiento, Kurzweil (2012) propone que el neocórtex opera bajo un principio básico, el *Pattern Recognition Theory of Mind*, que podría replicarse en soportes artificiales para recrear procesos semejantes a los humanos. Entender por máquina un sistema computacional que replica patrones neuronales no agota la discusión. Si tal dispositivo pudiera, además, emular la percepción sensorial y la interacción corporal con el entorno, como hipotetizan algunas perspectivas de la robótica avanzada (Bostrom, 2014) y de la inteligencia artificial encarnada (Kurzweil, 2012), estaríamos frente a una máquina dotada de un *cuerpo* artificial, pues tendrían sensores visuales, táctiles, auditivos, cinestésicos y motores que le permitirían situarse en un mundo y responder a él. Sin embargo, el desafío central radica en discernir si esta emulación de los sentidos equivale, en efecto, a la percepción vivida por un organismo biológico.

Partiendo de este contraste, la cuestión que se vuelve crucial en este punto es, aunque una máquina pudiera procesar información sensorial y generar respuestas funcionales coherentes, lo que permanece en entredicho es si ello conllevaría una *experiencia subjetiva*. Simular estímulos visuales mediante cámaras o reacciones afectivas mediante algoritmos no es lo mismo que *sentir* la emoción que despierta un paisaje o el dolor que provoca una herida. El cuerpo humano no solo percibe, sino que vive su percepción en continuidad con su memoria, su historia y sus

afectos, produciendo un impacto en su percepción del mundo y su interacción con él. Por ello, aunque la hipótesis de Dehaene (2014) o la de Kurzweil (2012) ofrecen un marco funcional para pensar una máquina como receptáculo de la conciencia, desde una perspectiva filosófica se puede sostener que la subjetividad no se reduce a la mera integración de información sensorial y computacional, sino que se arraiga en una corporeidad vivida que difícilmente puede ser replicada.

Un ejemplo particularmente ilustrativo de estas tensiones que se han relatado se puede encontrar a principios de los años 80's en la figura de Roy Batty, personaje del clásico cinematográfico *Blade Runner* (Scott, 1982). Precisamente en la escena cuando este replicante, un humanoide modelo Nexus-6, pronuncia su célebre monólogo: "I've seen things you people wouldn't believe. Attack ships on fire off the shoulder of Orion. I watched C-beams glitter in the dark near the Tannhäuser Gate. All those moments will be lost in time, like tears in rain..." (Scott, 1982, 1:45:23). Batty no solo enumera datos o imágenes que ha atestiguado como androide, sino que revela una subjetividad afectiva y consciente de la muerte, lo cual resulta extraordinario e inverosímil. Pues bien, este breve fragmento de la cinematografía funciona como una crítica anticipada al reduccionismo subyacente en las teorías que aspiran a transferir la conciencia a una máquina, pues aunque se lograsen emular patrones neuronales y procesos funcionales, la dimensión vivencial y corpórea de la conciencia podría no ser transferible, como se expone a continuación, en un debate que tiene como objetivo contrastar las teorías contemporáneas que sostienen que puede ser replicada y aquellas que afirman lo contrario.

La conciencia puesta en perspectiva

En el transcurso de las últimas tres décadas, los estudios científicos de la conciencia se ha centrado en identificar los correlatos neuronales y computacionales que posibilitan la experiencia consciente (Baars, 1988; Dehaene, 2014; Gazzaniga, 2018; Kurzweil, 2024; Tononi, 2008). Teorías que buscan desplazar las aproximaciones meramente especulativas en favor de marcos empíricamente verificables. Sus hipótesis, aunque diversas en sus mecanismos explicativos, comparten el objetivo de naturalizar la conciencia, o, dicho de otra manera, quieren explicarla como un fenómeno emergente a partir de procesos físicos y funcionales medibles en el cerebro.

Un primer modelo es el GNW, desarrollado por Dehaene (2014) a partir de la teoría original de Baars (1988), quien plantea que la conciencia emerge cuando la información es ampliamente distribuida e integrada en una red neuronal global que conecta regiones corticales frontales y parietales. Concretamente afirma

que la conciencia consiste en un *proceso de difusión global* de información en el cerebro. El cerebro humano habría desarrollado redes de largo alcance, especialmente en la corteza prefrontal, que permiten seleccionar información relevante y difundirla al conjunto del sistema nervioso. La función de la conciencia en este sentido tiene como objetivo mantener activa cierta información dentro de este espacio de difusión, haciéndola disponible para múltiples operaciones cognitivas, ya sea nombrarla, evaluarla, memorizarla o emplearla en la planificación de acciones futuras. “Computer simulations of neural networks show that the global neuronal workspace hypothesis generates precisely the signatures that we see in experimental brain recordings. It can also explain why vast amounts of knowledge remain inaccessible to our consciousness” (Dehaene, 2014, p. 182).¹ Esta actividad se correlaciona con ciertos marcadores fisiológicos, como la *onda P300*, la activación de neuronas piramidales y la sincronización oscilatoria de alta frecuencia. Sus aportaciones son significativas pues la GNW ha mostrado tener aplicaciones clínicas relevantes, como la detección de conciencia residual en pacientes en estados no comunicativos.

Por su parte, la IIT, formulada por Tononi (2008), sostiene que la conciencia no depende únicamente de la cantidad de información procesada, sino de la capacidad de un sistema para integrar esa información en una totalidad unificada. Según su teoría, un estado consciente se caracteriza por poseer un contenido rico y diferenciado, pero al mismo tiempo indivisible, de modo que la experiencia subjetiva surge precisamente de esta integración. Para cuantificarlo, Tononi introduce el *parámetro Φ (phi)*, que mide el grado de interconexión causal dentro de un sistema, cuanto mayor es Φ , mayor es el nivel de conciencia atribuible al sistema. En este sentido, la IIT no reduce la conciencia a correlatos neuronales locales, sino que la identifica con la forma en que la información se organiza y se mantiene integrada en un todo coherente.

De alguna manera, alineada con los esfuerzos realizados por la GNW, el atractivo de esta propuesta reside en su intento de ofrecer una formulación matemática y empíricamente contrastable de la conciencia, al punto que algunos experimentos con estimulación cerebral y electroencefalografía han tratado de estimar valores de Φ en sujetos humanos y animales. No obstante, esta teoría ha recibido críticas significativas porque su aplicabilidad práctica resulta

¹ “Las simulaciones computacionales de redes neuronales muestran que la hipótesis del espacio de trabajo neuronal global generan precisamente las señales que observamos en los registros cerebrales experimentales. También puede explicar por qué grandes cantidades de conocimiento permanecen inaccesibles a nuestra conciencia.” (Dehaene, 2014, p. 182) (Todas las traducciones son del autor).

limitada pues el cálculo exacto de Φ en sistemas complejos es prácticamente inabordable. Por otro lado, autores como Chalmers (1996) advierten que la IIT no distingue claramente entre la simulación de un proceso consciente y la existencia de una conciencia real. Si bien la IIT representa uno de los intentos más ambiciosos de formalizar la experiencia subjetiva, sigue siendo discutido hasta qué punto su formalismo logra captar lo que hace de la conciencia un fenómeno irreductible.

Por su parte, la propuesta WBE, impulsada por Kurzweil (2024) es, quizás, la más agresiva en afirmar la posibilidad de transferir la conciencia humana a una máquina. Esta corriente ligada al Transhumanismo sostiene que sería posible replicar digitalmente un cerebro humano mediante el escaneo detallado de sus estructuras sinápticas y funcionales. “A key issue I explored in *How to Create a Mind* is the philosophical and moral implications of replicating all the information in a human brain, which will be possible during the lifetimes of most people alive today” (Kurzweil, 2024, p. 117).² Bajo este supuesto, una copia funcional de la mente podría “correr” en un *hardware* alternativo. “At the stage of directly copying over the contents of living brains to nonbiological mediums, we transition from the merely simulated replicants I describe to actual mind uploading, also known as whole-brain emulation, or WBE” (p. 134).³ Kurzweil ha logrado posicionarse como uno de los más fervientes defensores de la idea de que la mente humana puede ser transferida a un soporte no biológico.

Este autor afirma que el cerebro es, en esencia, un sistema de reconocimiento de patrones jerárquico, es decir, un grupo de algoritmos diseñado para ejecutar diversas funciones, los cuales pueden ser descritos e implementados en diferentes sistemas computacionales de manera indistinta. En su más reciente obra, *The Singularity is Nearer* (2024), establece que, al capturar y replicar los patrones de activación y las conexiones neuronales, será posible reconstruir no solo las funciones cognitivas, sino también la experiencia subjetiva de un ser humano. Para el director de ingeniería de Google y el científico de mayor renombre en la actualidad especializado en Ciencias de la Computación e Inteligencia Artificial, el proceso de *mind uploading* no es una simple especulación futurista, sino una consecuencia natural del progreso tecnológico en neurociencia, informática y nanotecnología.

² “Un tema clave que exploré en ‘*How to Create a Mind*’ son las implicaciones filosóficas y morales de replicar toda la información en un cerebro humano, lo cual será posible durante la vida de la mayoría de las personas que viven actualmente” (Kurzweil, 2024, p. 117)

³ “En la etapa de copiar directamente el contenido de cerebros vivos a medios no biológicos, pasamos de los simples replicantes simulados que describo a la carga real de la mente, también conocida como emulación cerebral integral, o WBE” (p. 134)

La clave para comprender la perspectiva de Kurzweil (2024) es analizar su propuesta sobre el escaneo exhaustivo de las estructuras sinápticas y funcionales del cerebro humano, hasta un nivel tal que dicha práctica permita replicar el *mapa causal* de las interacciones neuronales. Para explicarlo de otra forma, el cerebro humano es un órgano complejo que genera conexiones físicas, las cuales facilitan la comunicación entre dos o más neuronas, o entre una neurona y una célula que trabajan juntas para permitir la cognición, el movimiento o la regulación de funciones vitales. Si se lograra realizar una copia exacta de este esquema funcional completo del cerebro, la mente, desde su perspectiva, podría ser replicada e implementada en un *hardware* alternativo. Como se puede observar, esta afirmación refleja la idea de que la continuidad de la identidad personal depende exclusivamente de la preservación de la dinámica causal en el cerebro, más que de su soporte biológico.

No obstante, Kurzweil (2002) reconoce también que este escenario abre profundos dilemas éticos y filosóficos que han sido debatidos desde los mismos diálogos platónicos. Y abiertamente se cuestiona sobre qué pasaría si existiera la posibilidad de “cargar” una mente en un soporte artificial, pues no estaríamos ciertos de lo que ocurriría con la noción de la identidad personal. “But is that person based on my mind file, who migrates across many computational substrates, and who outlives any particular thinking medium, really me? We come back to the same questions of consciousness and identity” (p. 110).⁴ Es importante hacer notar que el autor tampoco tiene una respuesta al dilema de qué surgiría si una copia exacta del cerebro correspondería al mismo individuo o a un nuevo ente consciente. Él mismo señala en sus apuntes que este proceso sería uno de los cambios más radicales en la historia de la humanidad, pues pondría en cuestión los límites entre lo humano y lo artificial, lo natural y lo sintético. Más allá, implicaría que la mortalidad podría ser superada, al transferir la mente a sistemas potencialmente inmortales y mejorables.

Bostrom (2014), uno de los primeros filósofos contemporáneos en abordar los riesgos de la *superinteligencia* y el consecuencialismo, reconoce al WBE como una de las vías más plausibles hacia la creación de inteligencias artificiales con capacidades comparables y eventualmente superiores a las humanas. “WBE is one possible path to superintelligence. Once we have the ability to emulate human brains, we could enhance them by running them faster, connecting them together,

⁴ “¿Pero esa persona que está basada en mi archivo mental, que migra a través de muchos sustratos computacionales y que sobrevive a cualquier medio de pensamiento particular, es realmente yo? Volvemos a las mismas preguntas sobre la conciencia y la identidad” (Kurzweil, 2002, p. 110)

or modifying them in other ways” (p. 36).⁵ Incluso detalla un mapa de ruta técnico que va desde el desarrollo de mejores técnicas de neuroimagen hasta la construcción de supercomputadoras capaces de correr tales simulaciones. “If whole brain emulation becomes possible, it will be scalable. One could, in principle, create a large population of emulations, all based on the same original brain or on different ones. The economic and social consequences would be profound” (p. 33).⁶ Sin embargo, es mucho más cauteloso y escéptico que Kurzweil (2024) al advertir que la *emulación cerebral completa* no solo abriría la posibilidad de replicar mentes humanas, sino también de multiplicarlas masivamente, lo cual podría desencadenar dinámicas sociales, económicas y políticas sin precedentes, además de problemas éticos sobre la explotación o el sufrimiento digital.

Así, mientras Kurzweil (2024) proyecta el WBE como una promesa de continuidad de la mente y de superación de los límites biológicos, Bostrom (2014) subraya tanto sus potenciales beneficios como sus riesgos. El primero lo presenta como la vía más directa hacia la inmortalidad tecnológica y el advenimiento de la *singularidad*; el segundo, como una de las rutas posibles hacia la *superinteligencia*, pero con serias implicaciones de seguridad existencial para la humanidad. Ampliar el análisis del WBE permite situar al lector en el corazón del debate y cuestionarse si estamos ante una utopía de trascendencia tecnológica o frente a una posibilidad cargada de dilemas ontológicos, éticos y políticos de gran escala.

Las críticas al reduccionismo materialista

A pesar de sus contrastes y distintos puntos de partida, las tres teorías expuestas, la GNW, la IIT y la WBE, comparten un enfoque funcionalista y fisicalista. Exploradas a través de un sentido crítico, podría afirmarse que pecan de ser reduccionistas al concebir la conciencia exclusivamente como un producto emergente de la actividad neuronal y, por tanto, potencialmente modelable en cualquier soporte computacional. Todas ellas enfrentan dificultades para explicar lo que la filosofía de la mente denomina *conciencia fenoménica* (Chalmers, 1996; Nagel, 1974), es decir, el *qué se siente* al tener una experiencia. Esta dimensión subjetiva, vivida, encarnada y afectiva es el núcleo de muchas de las críticas dirigidas al reduccionismo materialista y a los intentos de replicar computacionalmente la

⁵ “WBE es una posible vía hacia la superinteligencia. Una vez que tengamos la capacidad de emular cerebros humanos, podríamos mejorarlos haciéndolos funcionar más rápido, conectándolos entre sí o modificándolos de otras maneras” (Bostrom, 2014, p. 36)

⁶ “Si la emulación del cerebro completo se vuelve posible, será escalable. Uno podría, en principio, crear una gran población de emulaciones, todas basadas en el mismo cerebro original o en diferentes. Las consecuencias económicas y sociales serían profundas” (2014, p. 33)

mente humana. Es precisamente en este punto donde la fenomenología, como corriente dentro de la filosofía, introduce elementos clave al subrayar que la conciencia no es una función abstracta ni un mero procesamiento de información, sino una vivencia situada en un cuerpo, en un mundo, y en una historia personal (Merleau-Ponty, 1993). Esta perspectiva invita a poner en cuestión la posibilidad misma de trasladar o copiar la conciencia humana a una máquina sin pérdida esencial.

Filósofos y expertos en neurociencias como Chalmers (1996), Block (2007) y Gazzaniga (2018) han advertido que la reproducción funcional del cerebro no garantiza la preservación del yo subjetivo ni la emergencia de una experiencia consciente auténtica. Ello se debería principalmente a que la identidad personal no depende únicamente de un patrón funcional, sino de la continuidad vivida del cuerpo en el tiempo, de su anclaje en la historia encarnada del sujeto y de su inserción en un mundo intersubjetivo. Chalmers (1996) formula una distinción entre los “problemas fáciles” y el “problema difícil” de la conciencia. Los primeros se refieren a funciones cognitivas explicables mediante la neurociencia como la atención, la memoria o el procesamiento sensorial; el segundo se refiere a por qué y cómo ciertas actividades cerebrales dan lugar a la experiencia subjetiva, es decir, por qué existe “algo que se siente al estar consciente”. Argumenta que la conciencia fenoménica no puede explicarse completamente en términos físicos o funcionales y propone que debe considerarse un rasgo fundamental de la realidad, tal como el espacio o la masa. Esta posición, conocida como dualismo naturalista, desafía las teorías que buscan emular la mente humana sin abordar su dimensión cualitativa. En el contexto de la transferencia de conciencia, el “problema difícil” implica que, aunque se pudiera copiar la funcionalidad cerebral, ello no garantiza que emerja una experiencia subjetiva auténtica en el soporte digital.

Aunque Chalmers (1996) intentó durante mucho tiempo evitar el dualismo mente-cuerpo por ser una postura difícil de sostener en el marco de la ciencia contemporánea, tras un profundo análisis concluye que el dualismo no solo es una opción plausible, sino también intelectualmente satisfactoria. Su postura puede describirse como un “dualismo de propiedades” o “dualismo naturalista”, en el sentido de que considera a la conciencia como una propiedad fundamental de la realidad, no reducible a procesos físicos. Sin embargo, también sostiene que la conciencia podría surgir en cualquier sistema que reproduzca la organización causal adecuada, lo cual abre la posibilidad de su emergencia en soportes no biológicos basado en el principio de invariancia organizacional: “I am happy to embrace the conclusion that if cognitive dynamics are computable, then the right

sort of computational organization will give rise to consciousness” (p. 322).⁷ Sin embargo, la posibilidad de describir en términos computacionales la organización causal del cerebro plantea dificultades técnicas y teóricas aún no resueltas. Este es precisamente el dilema insalvable, al menos en el estado actual del conocimiento. La noción de que la organización causal del cerebro, la manera en que sus partes interactúan para producir estados conscientes, pueda describirse en términos de una abstracción computacional, sin necesidad de replicar la materia biológica es controvertida. En teoría, no sería necesario copiar un cerebro neurona por neurona, sino capturar su estructura causal relevante para poder ser implementada en otro soporte.

Pero, como el propio Chalmers (1996) afirma, esto supone que la organización relevante del cerebro pueda describirse a partir de un número finito de partes, cada una con un número también finito de estados posibles. Sin embargo, no todos aceptarían esta premisa. Algunos podrían objetar que, para reflejar adecuadamente el papel esencial del procesamiento continuo en una neurona, sería necesario admitir un número infinito de estados. Y aún más, hay quienes sostendrían que las transiciones entre esos estados infinitos podrían ser, en principio, no computables. Esta dificultad marca los límites epistemológicos y técnicos que enfrenta cualquier intento de replicar la conciencia a través de la emulación computacional. En tal caso, es importante distinguir entre un dualismo ingenuo, como el que subyace a ciertas propuestas transhumanistas que conciben la mente como un *software* que podría descargarse en cualquier *hardware*, y un dualismo más matizado, como el propuesto por Chalmers (1996), quien sostiene que la conciencia es una propiedad fundamental no reducible a lo físico, aunque, en principio, podría surgir en sistemas que reproduzcan la organización causal adecuada.

Para sostener su argumentación, Chalmers (1996) retoma la propuesta de Nagel (1974), quien introduciría el concepto “qué se siente” (*what it is like?*) como núcleo de la conciencia. Este autor sostiene que una mente consciente se caracteriza por tener una perspectiva interna, una vivencia subjetiva inaccesible desde la tercera persona. Aunque se pueda conocer todos los datos objetivos sobre un murciélago, nunca sabremos cómo se siente ser uno, porque eso implica una experiencia encarnada desde dentro. Este argumento establece un límite epistemológico a la comprensión de la conciencia desde los modelos computacionales. Aplicado al debate sobre su transferencia, implica que no hay forma de verificar si

⁷ “Estoy contento de aceptar la conclusión de que si la dinámica cognitiva es computable, entonces el tipo adecuado de organización computacional dará lugar a la conciencia” (Chalmers, 1996, p. 322)

una computadora siente algo, incluso si replica a la perfección las funciones cerebrales humanas. El aspecto cualitativo de la conciencia, es decir, el “qué se siente”, constituye un límite intrínseco, una barrera ontológica y epistemológica, dado que la naturaleza subjetiva de la experiencia consciente no es accesible desde una perspectiva externa ni reproducible a través de una simulación computacional.

Por su parte, Block (1995) añadirá un elemento más a este debate al hacer la distinción entre *A-consciousness* ‘conciencia de acceso’ y *P-consciousness* ‘conciencia fenoménica’. La primera se refiere a la disponibilidad de contenidos mentales para el razonamiento, la toma de decisiones y la acción verbal; la segunda, a las cualidades subjetivas de la experiencia. Block (2007), critica las teorías que equiparan conciencia con funcionalidad, argumentando que es posible tener conciencia fenoménica sin conciencia de acceso, como ocurre en ciertos estados perceptivos preatencionales o en sueños lúcidos. “P-conscious content is phenomenal, whereas A-conscious content is representational. It is of the essence of A-conscious content to play a role in reasoning, and only representational content can figure in reasoning. [...] many phenomenal contents are also representational” (p. 170).⁸ Distinguir entre acceso funcional y experiencia fenoménica es esencial para comprender los límites de las propuestas de inteligencia artificial puesto que una máquina puede operar datos sin que ello conlleve una vivencia consciente, careciendo por tanto de una experiencia fenoménica.

Más recientemente, Gazzaniga (2018), subraya que la conciencia no es un módulo aislado ni una función central, sino un instinto complejo que emerge de la interacción dinámica entre circuitos y capas distribuidos entre el cerebro primitivo y el neocórtex. Desde su punto de vista, todos compartimos ese estado que llamamos conciencia como una forma de estar atentos al fluir de nuestros pensamientos, deseos, emociones y sentimientos acerca del mundo, de los otros y de nosotros mismos. Se trata de una vivencia que, además de estar siempre presente, es profundamente personal, nos define y establece los límites de nuestra experiencia. En última instancia, es lo que da forma al hecho mismo de vivir. “The conscious self seems to ride above the physical brain and all its layers and modules. It seems that without it, we would be nothing but one of the automatons that Descartes observed [...]. A machine” (p. 174).⁹ Y desde esta perspectiva, apunta que cualquier intento de

⁸ “El contenido P-consciente es fenomenal, mientras que el contenido A-consciente es representacional. Es parte de la esencia del contenido A-consciente desempeñar un papel en el razonamiento, y solo el contenido representacional puede participar en el razonamiento. [...] muchos contenidos fenomenales también son representacionales” (Block, 2007, p. 170)

⁹ “El yo consciente parece cabalgar por encima del cerebro físico y todas sus capas y módulos. Parece que sin él, no seríamos más que uno de los autómatas que Descartes observó [...]. Una máquina” (Gazzaniga, 2018, p. 174)

replicar la conciencia a partir de una metáfora computacional clásica que concibe al cerebro como una máquina estaría condenado a una comprensión parcial del fenómeno.

En esta misma línea de ideas, Chalmers (1996) ya había advertido sobre un riesgo inherente de los modelos computacionales de la conciencia al relatar el desarrollo de sistemas que actúan como si fueran conscientes, pero que carecen de una experiencia interna genuina. Estos sistemas, denominados *zombis*, podrían exhibir conductas sofisticadas, incluso, emular emociones y lenguaje, sin poseer conciencia fenoménica. “According to these objections, computational systems would at best have the hollow shell of a mind: they would be silicon versions of a zombie” (p. 314).¹⁰ La conciencia no es reducible a sus correlatos neuronales ni funcionales, y cualquier intento de transferencia plena que omita esta dimensión está condenado a ser, en el mejor de los casos, una simulación sin sujeto, un *zombi*. Este es el punto de partida para comprender los límites ontológicos y epistemológicos de la emulación digital de la mente humana. Todos estos enfoques coinciden en que la experiencia consciente posee un carácter subjetivo, encarnado y cualitativo que escapa a los modelos computacionales puramente funcionales (Varela *et al.*, 2016). De ahí que la conciencia no sea reducible a correlatos neuronales ni a descripciones de acceso funcional dado que una reproducción de funciones no equivale a una transferencia de la experiencia vivida.

En este punto resulta crucial incorporar la perspectiva filosófica de la fenomenología, particularmente la formulación hecha por Merleau-Ponty (1993). Aun cuando este escribió su obra décadas antes del auge de la neurociencia computacional y de la inteligencia artificial, su propuesta ofrece herramientas conceptuales valiosas para pensar críticamente el rol de la conciencia en el contexto actual. Desde la perspectiva de la fenomenología, la conciencia no es una operación abstracta ni una interfaz de procesamiento de datos, sino una apertura encarnada al mundo, donde el cuerpo vivido, la afectividad y la temporalidad juegan un rol constitutivo. La conciencia, entonces, es inseparable del cuerpo, entendido no como objeto anatómico, sino como estructura de sentido y presencia en el mundo. Esta concepción desafía los intentos contemporáneos de disociar la mente de su corporeidad, tal como se observa en las corrientes del funcionalismo o del Transhumanismo. Así, la fenomenología no responde directamente a esos

¹⁰ “Según estas objeciones, los sistemas computacionales, como máximo, tendrían la cáscara vacía de una mente: serían versiones de silicio de un zombi” (Chalmers, 1996, p. 314)

modelos, pero sí antecede y enriquece su crítica, recordando que la experiencia consciente está siempre situada, encarnada y mediada históricamente. El cuerpo no es un mero soporte físico ni un simple *hardware* para un *software*, es, por el contrario, el centro activo desde el cual se despliega la percepción y la experiencia. “El cuerpo es nuestro medio general de poseer un mundo. Ora se limita a los gestos necesarios para la conservación de la vida y, correlativamente, propone a nuestro alrededor un mundo biológico” (Merleau-Ponty, 1993, p. 163). El cuerpo no tiene una conciencia, es conciencia, en tanto que vive, actúa y percibe. Esta idea desafía radicalmente cualquier proyecto de desanclar la mente de su base biológica. La conciencia humana no puede trasladarse a un soporte digital sin perder su anclaje fenomenológico; es decir, su modo concreto de estar en el mundo a través del cuerpo vivido.

Colombetti (2014) ha profundizado en esta misma línea de pensamiento, mostrando que la conciencia está afectivamente encarnada y sosteniendo que las emociones no son eventos mentales que ocurren en un sistema cognitivo, sino modos corporales de relación con el mundo. La tristeza, la alegría o el miedo no son datos que el cerebro procesa, sino vivencias que se sienten en el cuerpo; un nudo en la garganta, una presión en el pecho, un estremecimiento en la piel. Para ella, la mente se genera o se hace presente a partir del organismo vivo, gracias a su organización específica y a la manera en que este se relaciona con el mundo. “The body of the enactive mind is thus not just the perceiving and acting body but the living body, and as such it includes, for example, the viscera, the circulatory system, the immune system, and the endocrine system” (p. xv).¹¹ La conciencia afectiva y corporal no podría separarse del cuerpo humano sin reducirla a una caricatura funcional; no se trata de datos procesados en un sistema, son experiencias sentidas que se manifiestan en el cuerpo. Este enfoque también problematiza las ideas de transferencia de conciencia al señalar que un sistema que no tiene un cuerpo biológico que sienta y padezca, no puede tener experiencias afectivas reales, solo simulaciones externas de emociones, simulaciones conductuales de afectos.

Desde la ciencia cognitiva contemporánea, la teoría de la cognición encarnada o *embodied cognition*, sostiene que la mente no puede ser considerada independiente del cuerpo ni de su interacción sensomotriz con el entorno. El enfoque enactivista, una extensión de esta teoría, sostiene que el cuerpo vivo es un sistema que

¹¹ “El cuerpo de la mente enactiva no es solo el cuerpo que percibe y actúa, sino el cuerpo viviente, y como tal incluye, por ejemplo, las vísceras, el sistema circulatorio, el sistema inmunológico y el sistema endocrino” (Colombetti, 2014, p. xv)

se autoproduce y se automantiene, y que, a partir de esa dinámica, hace emerger significados. La mente no es un cerebro que representa el mundo, sino un cuerpo vivo que “enacta” (trae a la mano) un mundo de sentido acoplado con su entorno. Los procesos cognitivos, por tanto, pertenecen al ámbito relacional entre el organismo y su entorno. “Today, this idea of cognition as based on modal sensorimotor processes is central to the approach called “grounded cognition,” where “grounded” means based on body states, situated action, and modal perception-action systems” (Varela *et al.*, 2016, p. xxv).¹²

De modo que percepción, pensamiento y acción están profundamente entrelazados con la corporalidad. Para autores como Varela *et al.* (2016) y Nöe (2004), el conocimiento emerge de esta relación dinámica entre el cuerpo y el mundo que lo rodea. Y los modelos computacionales clásicos, los cuales están basados en esquemas *entrada-procesamiento-salida*, son criticados precisamente por omitir esta dimensión constitutiva. La visión enactivista critica este modelo al afirmar que el conocimiento y la experiencia no pueden surgir en un sistema que no interactúa activamente con el mundo físico mediante un cuerpo que siente, se mueve y responde. Si se desliga la conciencia del cuerpo y del mundo externo, no queda una mente consciente, sino una abstracción funcional sin vivencia.

En consecuencia, cualquier intento de trasladar la conciencia humana a un soporte digital que carezca de interacción sensorial y motriz con el mundo estaría condenado, como mínimo, a producir una simulación funcional sin vivencia subjetiva. Es así como Nöe (2004) afirma que la percepción no es algo que ocurre exclusivamente en la cabeza o en el cerebro, sino en el conjunto dinámico del cuerpo en interacción con su entorno. De tal modo que la conciencia puede considerarse intrínsecamente corporal y no un *software* que puede “migrarse” de una plataforma biológica a otra digital sin perder su esencia. Afirma que el cuerpo no es un contenedor pasivo de la mente sino aquella condición esencial que lo posibilita, de tal modo que se puede argumentar que los intentos de transferir la conciencia prescindiendo del cuerpo como instancia vivida están condenados a fracasar en su aspiración de preservar la experiencia subjetiva plena del hombre. Si bien es posible que ciertos aspectos funcionales de la conciencia pudieran simularse, la vivencia encarnada, afectiva y situada que constituye nuestra conciencia humana no sería transferida íntegramente a una máquina.

¹² “Hoy en día, esta idea de la cognición basada en procesos sensorimotores modales es central en el enfoque llamado ‘cognición fundamentada’, donde ‘fundamentada’ significa basada en los estados corporales, la acción situada y los sistemas de percepción-acción modales” (Varela *et al.*, 2016, p. xxv)

Sobre la posibilidad real de transferir la conciencia

Como ya se ha planteado, los avances tecnológicos de este siglo han renovado el interés en la posibilidad de trasladar aspectos de la mente humana a sistemas artificiales. Sin embargo, como también se ha podido observar, las dimensiones filosófica y fenomenológica muestran que esta posibilidad no está exenta de profundas limitaciones conceptuales. Entonces, lo que corresponde es analizar cuáles son los alcances reales de lo que podría transferirse a una máquina con la tecnología actual, y cuáles son los límites estructurales de la conciencia vivida.

Desde un punto de vista funcional, autores como Block (2007) afirman que es factible que ciertas capacidades cognitivas humanas puedan ser replicadas o incluso superadas por sistemas artificiales. Procesos como la atención focalizada, la memoria operativa, la toma de decisiones basada en algoritmos, el aprendizaje adaptativo mediante técnicas de *machine learning* o el procesamiento del lenguaje natural muestran en la actualidad resultados notables en el ámbito de la inteligencia artificial (Block, 2007). Estas funciones corresponden a lo que se denomina *conciencia de acceso*, refiriéndose al procesamiento de información disponible para guiar la acción, el pensamiento y la conducta observable. En este sentido, sí es posible construir sistemas que simulen aspectos externos del comportamiento consciente, generando la apariencia de agencia o incluso de diálogo significativo.

En esta misma línea de ideas, sería posible construir sistemas que imiten de manera convincente ciertos aspectos del comportamiento consciente, generando la apariencia de agencia. Sin embargo, tal simulación no implica la presencia de una conciencia fenoménica y no garantiza que exista experiencia subjetiva tras esos procesos (Chalmers, 1996). Para ilustrar esta distinción, puede considerarse el caso de un vehículo autónomo, un sistema de inteligencia artificial que puede ser programado para conducir con gran eficacia, optimizando rutas, respondiendo a imprevistos e interactuando con otros vehículos en tiempo real. Pero que un automóvil se conduzca de manera segura no significa que experimente la conducción como una vivencia, ni que se sitúe en el mundo como un sujeto encarnado. Esta máquina no siente ansiedad en condiciones climáticas adversas, ni placer en la conducción fluida, es decir, no tiene una vivencia subjetiva incorporada en el acto que ejecuta. Por tanto, la diferencia entre actuar como si se fuera consciente y serlo realmente no radica únicamente en la presencia de *qualia* o contenido fenomenológico, sino en la existencia de un cuerpo consciente que viva dichas experiencias, como lo señala la tradición fenomenológica.

Aunque algunos autores, como Kurzweil (2024), sostengan que la emulación completa del cerebro humano podría eventualmente generar un sistema dotado

de subjetividad, otros autores como Chalmers (1996), el mismo Block (2007) o Gazzaniga (2018) subrayan que no existe un marco conceptual ni empírico que permita *verificar* si un sistema artificial siente o simplemente simula sentir. Desde esta perspectiva, la figura del replicante Roy Batty en *Blade Runner* opera como una metáfora contundente: un ser artificial que pretende expresar recuerdos, afectividad y conciencia de la muerte, pero que al final del día se trata simplemente de ciencia ficción. La capacidad del androide de enunciar estas experiencias en un discurso no garantiza que tales vivencias tengan un correlato fenoménico real. Mientras los sistemas artificiales actuales no cuenten con un cuerpo vivido ni con una historia afectiva encarnada, difícilmente podrán generar una conciencia comparable a la humana.

Es preciso insistir aquí en que la conciencia humana no puede reducirse a un conjunto de funciones ejecutables, sino que se manifiesta como una construcción narrativa en el tiempo, un todo de manera integrada. Bajo la perspectiva de Gazzaniga (2018), la conciencia permite al sujeto orientarse hacia el mundo, integrar sus vivencias y conferirles coherencia desde una perspectiva situada y afectiva. Si bien un sistema artificial podría ser programado para construir un relato de sí mismo, esto no basta para concluir que haya conciencia en sentido fenomenológico, pues la narratividad por sí sola no implica una experiencia sentida. Desde la perspectiva inspirada por Merleau-Ponty (1996) y las aproximaciones contemporáneas de autores como Colombetti (2014), la conciencia está enraizada en la historicidad del cuerpo vivido y en su afectividad, es decir, no solo en el relato que se construye, sino en la manera encarnada en que ese relato se siente y se vive. Transferir información no garantiza la continuidad del yo, pues no puede asegurarse que la conciencia que emerge en un nuevo soporte pueda ser la misma persona y no una copia sin conciencia del ser original. Así, aunque una inteligencia artificial afirmara ser consciente y actuara como tal, no tendríamos medios confiables para verificar si hay experiencia subjetiva tras sus afirmaciones. Este es el dilema precisamente que plantea Chalmers (1996) a través de la idea del zombi, una entidad que se comporta como si tuviera conciencia, pero que carece completamente de experiencia interna. Así pues, aunque ciertos aspectos funcionales de la mente puedan replicarse artificialmente, no implica que la conciencia fenomenológica emerja con ello. La posibilidad de que un ente como Roy Batty posea una subjetividad plena permanece, por ahora, como una conjetura filosófica abierta más que como una certeza empírica.

Esta alegoría de Batty sirve para ejemplificar claramente que la conciencia no es solo un procesamiento eficiente de datos, sino una experiencia trágica, sensible

y encarnada. Pretender transferir esa *totalidad* a una máquina es una ilusión que subestima la profundidad de lo que significa ser humano. Por consiguiente, es posible afirmar que la conciencia como experiencia vivida permanece fuera del alcance de los modelos computacionales. La transferencia funcional puede ser útil para la construcción de agentes artificiales eficaces, pero no equivale a trasladar el yo humano. Esta distinción entre simular e incorporar, entre representar y vivir, es esencial para comprender los límites filosóficos y fenomenológicos de la transferencia de conciencia. Las tensiones entre los avances técnicos en inteligencia artificial y las limitaciones impuestas por la filosofía de la mente y la fenomenología no son meramente teóricas. Más allá de los límites conceptuales sobre la posibilidad de transferir la conciencia humana, este debate plantea considerar los problemas filosóficos de fondo que pueden afectar también nuestra comprensión de la libertad, la identidad personal, la responsabilidad moral y la dignidad humana, los cuales podrán ser tratados en futuras disertaciones e investigaciones.

El objetivo de este artículo ha sido exponer que las teorías que sostienen la posibilidad de emulación de la mente suelen partir de una visión reduccionista, frecuentemente fisicalista, según la cual todo estado mental es reducible a un estado físico o funcional del cerebro. Esta postura, aunque útil para ciertos fines tecnológicos, omite la dimensión subjetiva y encarnada de la experiencia misma. Reducir la mente a un conjunto de algoritmos equivale a ignorar la vida consciente en su densidad existencial, y corre el riesgo de cosificar al ser humano, tratándolo como un objeto técnico más. La conciencia no es una función entre otras, es el modo en que el ser humano habita el mundo y tratarla como un módulo fácilmente replicable es reducir lo irreducible.

Paradójicamente, muchas de las propuestas que se presentan como post-cartesianas reproducen, en la práctica, un dualismo renovado en donde la mente funciona como *software* y el cuerpo es un *hardware* del que se puede prescindir. De acuerdo con el mismo Bostrom (2014), este nuevo dualismo cuerpo/datos o carne/código posibilita visiones fragmentadas donde el cuerpo biológico se percibe como obsoleto frente a un soporte artificial idealizado y mejorado. Aunque estas corrientes de pensamiento afirmen que mente y cuerpo están unidos, sus modelos operan contradictoriamente bajo la lógica de que la mente puede separarse del cuerpo y ejecutarse en otra base material. Esta perspectiva ignora que la conciencia no es un módulo intercambiable, sino un fenómeno relacional y situado, anclado en un cuerpo vivido. Además, resulta peligrosa, pues se introduce un nuevo tipo de dualismo entre cuerpo biológico (prescindible) y un soporte artificial (mejorado) aspiracional, el cual alimenta visiones donde la materialidad orgánica

es vista como obsoleta o inferior, desplazable y de la que podemos prescindir. Este enfoque termina por revalorizar entonces al cuerpo artificial y desvalorizar el cuerpo vivido, ignorando que la conciencia no es un *software* migrable, sino un fenómeno relacional y situado.

Los intentos de trasladar la conciencia humana a una máquina generan debates profundos sobre el reduccionismo de la experiencia, los dualismos implícitos, una crisis de la autonomía y la desestabilización de la identidad personal. Todos estos problemas no pueden resolverse desde una lógica puramente técnica, pues afectan la autocomprensión filosófica de lo que significa ser un sujeto. Para Chalmers (1996), una de las claves para abordar estos dilemas reside en distinguir con claridad entre la simulación funcional de procesos mentales y la presencia efectiva de una conciencia vivida.

El análisis hasta aquí desarrollado permite sostener que la idea de transferir la conciencia humana a una computadora, si bien puede ser seductora desde la lógica del progreso tecnológico, también encuentra límites profundos en el plano filosófico, fenomenológico y ontológico. “Present-day scientific theories hardly touch the really difficult questions about consciousness. We do not just lack a detailed theory; we are entirely in the dark about how consciousness fits into the natural order.” (Chalmers, 1996, p. xi).¹³ La tecnología puede emular funciones cognitivas, sí, pero no garantiza la emergencia de una subjetividad. La diferencia entre hacer “como si” y experimentar realmente permanece irresuelto. La conciencia no flota en un vacío, ni puede ser desanclada sin pérdida, pues está entrelazada de sentidos, emociones, percepciones y movimientos que se viven desde un cuerpo propio. Por ello, la reducción de la mente a datos trasladables implica una visión empobrecida del *ser humano*, en la que desaparece la vivencia, la fragilidad y la historicidad de la experiencia personal. Por ello, se puede afirmar que la conciencia es inseparable de la corporeidad.

Conclusiones

Frente a las expectativas que proyecta la posibilidad de una conciencia humana trasladable a sistemas artificiales, los análisis filosóficos y fenomenológicos aquí revisados permiten sostener una posición más crítica y matizada. Esta comprensión exige el diálogo entre diferentes disciplinas. Las neurociencias, la filosofía, la psicología, la antropología y la ética deben converger para repensar

¹³ “Las teorías científicas actuales apenas abordan las preguntas realmente difíciles sobre la conciencia. No solo carecemos de una teoría detallada; estamos completamente en la oscuridad sobre cómo la conciencia encaja en el orden natural” (Chalmers, 1996, p. xi)

qué entendemos por mente, por sujeto, por identidad. Es oportuno señalar que aceptar que la conciencia no es plenamente transferible no implica un rechazo total al progreso, sino una defensa del respeto por la complejidad de lo que significa ser humano. La técnica puede y debe contribuir siempre a nuestro bienestar, pero no puede sustituir el enigma de la experiencia vivida ni convertirnos en artefactos de nosotros mismos.

El recorrido realizado permite concluir que, si bien la tecnología contemporánea ha logrado avances notables en la simulación de funciones cognitivas como la atención, la memoria o el lenguaje, la transferencia plena de la conciencia humana permanece fuera del alcance conceptual, ontológico y existencial de estos desarrollos. Y también se puede concluir que la conciencia no se reduce a un conjunto de funciones ni a un flujo de información porque se trata de una experiencia vivida, encarnada, situada y afectiva que emerge de la interacción dinámica entre cuerpo, mundo y temporalidad que da pie a dos factores críticos, la intencionalidad del sujeto ante el mundo y su subjetividad. Intentar capturar esta experiencia en un soporte artificial equivale a confundir la función con el fenómeno, la simulación con la vivencia. Los sistemas artificiales actuales, por sofisticados que sean, pueden reproducir patrones de comportamiento, pero no generar subjetividad consciente.

Como plantea Gazzaniga (2018), la conciencia es un fenómeno distribuido y profundamente enraizado en la arquitectura del cerebro como un instinto complejo. Comprenderla exige superar las metáforas computacionales reductivas y avanzar hacia una teoría que integre, de manera rigurosa, neurociencia, filosofía, fenomenología y ciencias cognitivas. Gazzaniga, director del Centro de Neurociencia Cognitiva SAGE, cuya labor ha logrado avances importantes en la comprensión de la lateralización funcional en el cerebro humano, concluye que los humanos nunca construiremos un artefacto que imite nuestra conciencia personal porque las máquinas inanimadas basadas en silicio funcionan de una manera, mientras que los sistemas vivos basados en carbono lo hacen de otra. Para él, las primeras operan siguiendo un conjunto determinista de instrucciones, mientras que los segundos se valen de símbolos que implican siempre un cierto grado de incertidumbre. Desde su perspectiva, el esfuerzo humano por imitar la inteligencia y reproducir la conciencia en máquinas, objetivo recurrente en el campo de la inteligencia artificial, está destinado al fracaso.

Así pues, la hipótesis inicial queda confirmada: aunque ciertas funciones cognitivas pueden simularse, existen límites filosóficos y fenomenológicos que impiden una transferencia completa de la conciencia humana a una máquina. Estos límites no solo se remiten a la complejidad técnica, sino a la naturaleza misma de la expe-

riencia consciente, que implica un punto de vista interno, afectividad, historicidad y un cuerpo vivido, pero, sobre todo, subjetivo. Frente al entusiasmo por la inmortalidad digital, es necesario reafirmar el valor de la finitud, de la vulnerabilidad y de la historia encarnada que configuran nuestra humanidad.

En este sentido, la transferencia de la conciencia no es imposible por falta de tecnología, sino por un error de comprensión: se desconoce si es posible transferirla debido a que no se sabe del todo qué es. Reconocer este límite no es un fracaso, sino un acto de honestidad filosófica y una invitación a pensar más allá del algoritmo, desde y para la humanidad. No se trata de una falla técnica; el límite reside esencialmente en una comprensión insuficiente de qué es la conciencia. Como señala Chalmers (1996), mientras no resolvamos el “problema difícil”, todo intento de replicarla seguirá enfrentando barreras insalvables. Por ello, el desarrollo de una teoría integral de la conciencia que incorpore sus dimensiones fenomenológica, afectiva, corporal y subjetiva es hoy más necesario que nunca.

Referencias

- Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.
- Block, N. (2007). *Consciousness, function, and representation: Collected papers, vol. 1*. MIT Press.
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.
- Colombetti, G. (2014). *The feeling body. Affective science meets the enactive mind*. MIT Press.
- Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Viking.
- Ferrater, J. (1965). *Diccionario de Filosofía*. Editorial Sudamericana.
- Gazzaniga, M. S. (2018). *The consciousness instinct: unraveling the mystery of how the brain makes the mind*. Farrar, Straus and Giroux.
- Kurzweil, R. (2002). The evolution of mind in the twenty-first century. En J. W. Richards (Ed.), *Are we spiritual machines? Ray Kurzweil vs. the critics of strong A. I.* (pp. 95-125). Discovery Institute Press.
- Kurzweil, R. (2012). *How to create a mind. The secret of human thought revealed*. Viking.
- Kurzweil, R. (2024). *The singularity is nearer. When humans transcend biology*. Viking.

- Merleau-Ponty, M. (1993). *Fenomenología de la percepción*. Planeta-Agostini.
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435-450.
- Noë, A. (2004). *Action in Perception*. MIT Press.
- Scott, R. (Dir.). (1982). *Blade Runner* [Película]. The Ladd Company, Shaw Brothers, Blade Runner Partnership.
- Tononi, G. (2008). Consciousness as integrated information. A provisional manifesto. *The Biological Bulletin*, 215(3), 216-242. <https://doi.org/10.2307/25470707>
- Varela, F. J., Thompson, E. and Rosch, E. (2016). *The embodied mind: Cognitive science and human experience*. MIT Press.