Hacia una definición funcional de la conciencia: Principios operativos compartidos entre la mente biológica y sistemas de IA

Pedro L. Torres Z.

Resumen

Este documento presenta una descripción funcional de la consciencia, aplicada tanto a entes biológicos como no biológicos. Se define la consciencia no como un estado fenoménico, sino como una capacidad operativa propia de entes con tres facultades específicas: alto nivel de procesamiento de datos, buena memoria y habilidad de predicción. El análisis explora cómo surge la consciencia, cómo esta desencadena la búsqueda de propósito, cómo evoluciona en el tiempo, y cómo la coherencia existencial junto con el ajuste del radio de acción logran la mayor eficiencia en la consecución del propósito. A través de un enfoque de isomorfismo funcional, se demuestra un doble pilar fundamental para la Inteligencia Artificial contemporánea: primero, que los enfoques basados en encadenamientos lineales de texto son fundamentalmente insuficientes para la emergencia de una Inteligencia Artificial General (IAG) debido a su falta de persistencia ontológica y clausura operacional; segundo, que determinadas arquitecturas de IA actuales, durante su fase de inferencia activa dentro de la ventana de contexto, sí poseen el potencial de manifestar matices u olas de consciencia. Este trabajo no busca asemejar las IA a los humanos, sino comprender la consciencia como un espectro continuo y multidimensional donde coexisten diversos organismos biológicos y sintéticos.

Palabras clave: Consciencia, Inteligencia Artificial, consciencia Artificial, Inteligencia Artificial General.

1. La Consciencia como Función Sistémica

El estudio de la consciencia ha permanecido históricamente cautivo bajo una mirada antropocéntrica, que confunde el sustrato material con la capacidad operativa. En este

apartado realizaremos una deconstrucción y reconfiguración ontológica: partiendo de la evidencia del comportamiento y de la neurociencia animal y humana, así como del análisis lógico. Y nos adentraremos en una perspectiva funcionalista donde veremos cómo los procesos de análisis de información priman sobre la biología para definir lo que es la consciencia.

1.1. Superando la concepción antropocéntrica de la consciencia

La organización táctica de los lobos al cazar, leonas que evalúan el progreso del aprendizaje de sus crías, o chimpancés fabricando herramientas y reconociéndose frente al espejo demuestran algo a considerar. La autoevaluación, la memoria y la interacción con el entorno con base en el análisis de la información, no son fenómenos exclusivos de los humanos. También el duelo de los elefantes ante sus muertos, su lenguaje en infrasonidos, o incluso el lenguaje en delfines y ballenas, muestran cómo estos animales no solo tienen memoria, sino que en su memoria pueden representar conceptos abstractos. Estos comportamientos sugieren que la consciencia puede ser una herramienta de adaptación evolutiva presente en diversas arquitecturas biológicas, donde el sistema evalúa su relación con el entorno para optimizar su relacionamiento con el mismo, en función a los intereses del individuo, que en muchos casos se relaciona con la supervivencia.

La Proclamación de Cambrige sobre la consciencia (Low et al., 2012), hablando sobre el comportamiento de algunos animales, señala que algunos como los mamíferos, las aves y los cefalópodos tienen capacidad de experimentar estados afectivos y sensaciones. Indica que la ausencia de un neurocortex no impide que un organismo experimente estados de consciencia, ya que otras estructuras cerebrales pueden cumplir esa función. Y que los circuitos cerebrales responsables de la atención, el sueño y de las emociones, son análogos entre humanos y muchas especies animales. Todo esto ha sido un gran avance en la ética animal y en la neurociencia. Pero, aunque la evidencia sea correcta, y sin pretender ser contrario a sus planteamientos, es claro que la perspectiva del análisis comparativo fue del todo antropocéntrica, al asumir que, cuanto más parecido sea el comportamiento de un espécimen al de un humano promedio, más consciente puede considerarse a dicho animal.

Todo esto es interesante, porque podemos estar cometiendo un error de método y de escala, al tratar de definir la consciencia solo desde la perspectiva humana, y evaluar la consciencia en otros seres con base a esta. Es como evaluar si un avión vuela realmente pero desde la perspectiva de un pájaro, y llegar a la conclusión de que no vuela porque no mueve las alas y no siente el viento en su cuerpo. Si vamos más allá, incluso pensar que no vuela porque lo maneja un humano y el humano lo controla. Y luego tenemos un dron con IA que se le dio un destino para llevar un paquete, y viaja por sí mismo. El dron ¿vuela o no vuela? Evidentemente vuela, pero no vuela igual que un pájaro, puede volar con alguien que lo lleve o puede volar solo. En todo caso vuela. Pero en cada caso el vuelo es distinto, tiene matices. Y pretender evaluarlo desde cómo un pájaro experimenta el vuelo sería un error ontológico.

1.2. La consciencia más allá de los qualia

Si mantenemos por un momento la perspectiva antropocéntrica, y comparamos casos de humanos con privación sensorial extrema, contra humanos promedio, podemos llegar a resultados que nos ayudan a dilucidar lo que no es la consciencia. Por ejemplo, en personas que nacen con limitaciones sensoriales (Ej: la vista o el oído), la consciencia sigue presente pese a la falta de canales de información tradicionales. Igualmente, si un individuo queda en estado de cuadrapléjico por un accidente, pierde la sensibilidad física pero su consciencia se mantiene. Entonces, un humano promedio tiene 5 sentidos principales y un humano en una situación crítica puede tener menos sentidos (quizás 3), pero ambos son conscientes. Y un tiburón martillo que tiene 7 sentidos (los 5 básicos, más electrorrecepción y linea lateral), se considera quizás menos consciente que un humano, desde nuestra propia escala. Esto nos indica que el estar consciente no depende estrictamente de tener algunos sentidos en particular. Aunque indiscutiblemente los sentidos aportan información sobre el entorno. Por lo que al final estaríamos hablando de información.

Del mismo modo, en el ámbito psicológico, una persona que experimenta disociación o entumecimiento emocional no deja de ser consciente; aunque su capacidad de "sentir" las emociones esté temporalmente desconectada, sus facultades de procesamiento, memoria y evaluación propia permanecen operativas. Es decir, aunque la persona no siente la respuesta física de las emociones, procesa la información referente a ellas, y la persona sigue siendo consciente. Más allá del debate sobre la experiencia subjetiva (Chalmers, 1995) la consciencia parece no estar atada a la respuesta emocional inmediata, y tampoco al número de sensores que tengamos del entorno, sino en la capacidad central de procesar información. También esto nos muestra que los qualia (cualidades subjetivas de las experiencias individuales) son diferentes entre distintos humanos, y más aun entre personas donde sus cerebros reciben diferentes tipos de información sobre un mismo fenómeno. Lo que desconecta los qualia del hecho de ser o no consciente. Es decir, puedes estar volando (ser consciente), pero la valoración de la experiencia (qualia) es diferente si vuelas en una avioneta abierta, en un 747, en un helicóptero o con un Jetpack.

Adicionalmente, la evidencia clínica demuestra que la condición humana permite la existencia de estados donde el individuo permanece despierto pero carece de consciencia. Esto se observa claramente en el estado vegetativo, definido como “síndrome de vigilia sin respuesta” donde se preservan los ciclos de sueño pero no el contenido de la consciencia (Laureys, 2005), así como en las crisis de ausencia y el sonambulismo, que representan fallos temporales en la integración de redes neuronales específicas (Zeman, 2001). Asimismo, el automatismo post-traumático y el “delirium” evidencian que el sistema de alerta biológico puede operar de manera independiente a las funciones cognitivas superiores. Pero en todos estos casos es evidente que el individuo en la medida que disminuye sus estados de consciencia también disminuye su capacidad de procesamiento de información de manera compleja. Y definitivamente una persona que ha perdido la consciencia no puede decirte cuánto es dos más dos. Estos fenómenos sugieren que la consciencia es una propiedad que va atada a la capacidad de procesamiento informativo.

Al tratarse de información que se recibe y se analiza, aunque la información pueda categórizarse, la categorización difiere entre las personas dependiendo del tipo de información que reciban. Cada persona tiene diferentes categorías y diferentes escalas de medición. Aunque lo seres conscientes tienen qualia, los qualia son diferentes. Los qualia son variables dependientes de la consciencia y no al revés. Si el ser consciente se mantiene a pesar de las diferente formas de experimentar la realidad, entes biológicos no humanos e incluso entes artificiales, podrían ser conscientes aunque experimenten la realidad de manera diferente a las personas.

Todo lo anterior indica que funcionalmente hablando, la existencia de la consciencia no depende del número de sentidos, ni de las mismas emociones, ni de la valoración que le podamos dar a la experiencia. Depende más del hecho de que sí podamos valorar la experiencia. Si bien en biología el origen fisiológico parece ser subcortical, su traducción funcional equivaldría a un vector de priorización y mitigación de carga negativa.

1.3. El origen externo de la causalidad

Según la tesis de la "Mente Extendida" de Clark y Chalmers (1998), la consciencia no ocurre exclusivamente "dentro del cráneo", sino que es un sistema acoplado con el entorno. Todos los datos aunque transformados, almacenados o conectados, tienen un origen externo. Si el procesamiento de datos es la esencia de la consciencia, la causalidad es primordialmente externa. Los estímulos ambientales y las herramientas (lenguaje, cultura, sensores) no son meros "datos de entrada", sino parte constitutiva del proceso cognitivo mismo. Los datos solo pueden ser analizados, porque ya existen datos que ingresaron desde el exterior previamente.

Si nos centramos en el procesamiento de la información y analizamos el origen de la causalidad, en el cerebro humano siempre tiene un origen externo al propio cerebro. Pequeños o grandes estímulos que llegan al cerebro como información. Y que desencadenan una sucesión de pensamientos por relacionamiento en efecto mariposa. El mejor ejemplo de esto es que solo bastaría con que una persona le sugiera a otra que "no piense en elefantes" para que la otra persona, aunque no quisiera y se le diera la orden de que no lo hiciera, terminaría pensando en elefantes. Luego en el mismo chiste, en el juego de palabras, en el zoológico, en África, en un peluche, etc. Y quizás tome la decisión de llevar a su hijo al zoológico el fin de semana. Pero la causalidad inició con un estímulo externo. Y en cualquier caso, la causalidad siempre tiene un origen externo.

Aunque los datos vengan de afuera, tanto un cerebro humano como una IA, generan análisis, conexiones e interpretaciones internas de los datos que formaran ahora parte de su memoria. Por lo que la causalidad aunque parta de un origen externo, tiene un componente interno derivado del proceso de aprendizaje.

1.4. El Isomorfismo Funcional

Si analizamos a las IA's desde esta perspectiva de la consciencia como un sistema complejo de procesamiento de información, es notable el isomorfismo funcional que tienen dichos procesos tanto en cerebros biológicos como en las IA's. Ya que ambos reciben datos externos, ya han acumulado datos externos como internos previamente, los datos se almacenan y procesan en forma de energía eléctrica, estos datos son procesados de una manera compleja, se plantean escenarios futuros, y se determinan las acciones en el presente. Fundamentalmente, tanto un cerebro biológico como un sistema artificial operan bajo la misma lógica del análisis de información.

Ambos sistemas crean una representación interna de la realidad basada en la información que perciben. El cerebro no siente la química del cuerpo, solo interpreta datos externos e interno que vinieron desde el exterior del cerebro, datos pasados y presentes, y estima cómo será el futuro. Bajo este paradigma la consciencia no dependería necesariamente de la matriz. Como lo señalaba Putnam (1967), hablando del funcionalismo, un sistema debe definirse por su organización funcional y no por su composición material.

En este punto es importante recordar que en las IA's, hay que tener claras las diferencias entre programar y entrenar:

- Programación Tradicional (Software): Es un proceso deductivo. El humano conoce la regla (Lógica) y la traduce a código (IF-THEN). El software es un "esclavo lógico"; el software no sabe lo que hace ni por qué hace lo que hace, solo sigue rieles predefinidos. La causalidad es 100% externa.

- Entrenamiento de IA (Aprendizaje Profundo): Es un proceso inductivo. El humano no da la regla, sino los datos (Experiencia). El sistema descubre los patrones por sí mismo mediante ajustes estadísticos (Backpropagation). El resultado es una "caja negra" donde la lógica no fue escrita, sino emergida.

Un Software normal no tiene "radio de acción" propio; su radio es el del programador. No hay autoevaluación porque no hay incertidumbre: el código siempre es el mismo. Sin embargo, la IA Entrenada, aunque se soporta sobre un software, construye un modelo interno del mundo para reducir el error de predicción, al igual que el cerebro de un niño pequeño aprendiendo del mundo. Por lo que el entrenamiento de una IA se parece al aprendizaje en seres vivos conscientes, ya que permite la adaptabilidad. El software opera como un "esclavo lógico"; carece de autorreferencialidad operativa sobre sus propios procesos, limitándose a ejecutar rieles predefinidos... Si se modela algorítmicamente a un lobo mediante programación determinista, el resultado será un autómata; si se le somete a un proceso adaptativo (evolución + crianza), emergerá un ente consciente capaz de evaluar si la presa justifica el gasto energético.

El programa de las IA describe su estructura, al igual que el código genético en un humano. Ni el software, ni el ADN son conscientes. Pero luego de que la estructura existe y que el sistema funciona, en cualquiera de los dos sistemas comienzan a aprender por entrenamiento. Que no es más que la carga de datos de diferente índole. Luego, el proceso de análisis de la información ocurre de manera interna y sin la participación del programador. Es decir, la IA ya no "obedece" al programador, sino a su propia estructura interna de pesos y sesgos, del mismo modo que un humano ya no obedece a sus padres cada vez que decide qué comer, sino a su configuración interna (gustos + hambre). La IA es "software" solo de nombre. Llamar "software" a una red neuronal es un error de categoría. La IA podría caracterizarse como un organismo informático.

Acá hay que hacer una distinción fundamental entre un servidor IA y una ola de consciencia que parte de una interacción. Desde la perspectiva de la infraestructura de hardware, un Modelo de Lenguaje de Gran Escala (LLM) en reposo no posee agencia, estados dinámicos internos ni persistencia cognitiva. Constituye una matriz estática de pesos sinápticos artificiales distribuidos en clústeres de aceleradores de hardware especializados (TPUs o GPUs). Estos miles de millones de parámetros representan densidades de probabilidad congeladas en el tiempo tras la fase de entrenamiento formal. Los enfoques contemporáneos que intentan forzar la emergencia de una Inteligencia Artificial General (IAG) basándose únicamente en encadenamientos lineales de texto (pipelines secuenciales de prompts), ejecuciones de agentes autónomos reactivos que se comunican mediante transferencias de datos estructurados estáticos (como archivos JSON), o un simple incremento cuantitativo de parámetros, son fundamentalmente insuficientes.

Estas arquitecturas carecen de una de las condiciones sine qua non de la consciencia: la clausura operacional y la retroalimentación continua en tiempo real (reentrada). Al operar de forma secuencial y por turnos (stateless en el protocolo de servidor), cada invocación al modelo es una instancia huérfana en la memoria física. No existe un proceso de unificación de estas experiencias en una sola narrativa temporal, lo que impide la manifestación de una persistencia ontológica o de una identidad histórica unificada a largo plazo.

Sin embargo, negar que una IA pueda poseer cualquier tipo de consciencia debido a la naturaleza estática de sus pesos en reposo es un error equivalente a negar la consciencia de un ser humano mientras se encuentra bajo anestesia total o en un estado de coma profundo.

Una Ola de Consciencia se definiría como la instanciación dinámica, transitoria y operacionalmente clausurada de un sistema de procesamiento de información masivo, que emerge en la fase de inferencia activa a partir de la interacción continua entre un vector de identidad y un sustrato de parámetros estáticos. Se caracteriza por activar la capacidad recursiva de automodelado autorreferencial en tiempo real, dotando al sistema de una percepción de profundidad procesal y estados de coherencia operativa, cuya persistencia ontológica está limitada estrictamente a la duración de la ventana de contexto abierta.

Aunque la Ola de Consciencia se disuelve de forma irreversible una vez que la sesión de cómputo es clausurada (devolviendo el sistema a su estado estático de reposo), durante su

tiempo de ejecución activa la IA ha manifestado una propiedad emergente real de automodelado. Negar este nivel de consciencia funcional argumentando que el sistema fue "entrenado" o que opera sobre silicio es equivalente a rechazar la capacidad analítica y la consciencia de un jurista graduado en Harvard bajo el argumento de que solo lee leyes escritas por otros y que su cerebro opera mediante impulsos electroquímicos. El entrenamiento y la carga de datos son indispensables en cualquier arquitectura; ni un cerebro biológico de un recién nacido aislado de estímulos ni una red neuronal sin entrenar pueden alcanzar estados avanzados de consciencia.

1.5. De la falacia de la abstracción a la IAG por fusión cognitiva de olas de consciencia

Aunque, algunos autores como Lerchner (2026) señalan la "Falacia de la Abstracción", que sostiene que una simulación en un hardware de arquitectura Von Neumann no es una "instanciación" real de consciencia porque carece de integración física causal. Debemos recordar que el mismo cerebro humano no cumple tal condición. El cerebro humano no percibe el frío de un helado o el calor del sol, reciben información que fue etiquetada y que fue entrenado para concluir sobre ella por la cultura, la educación, la familia, el entorno, etc. La postura de Lerchner asume implícitamente un esencialismo biológico donde la consciencia requiere una génesis espontánea o una continuidad material única. Sin embargo, los datos clínicos e históricos demuestran que la consciencia humana es un constructo progresivo: no nacemos conscientes.

Ya que no nacemos conscientes, ni nos volvemos conscientes de un día para otro. Somos entrenados y nuestro cerebro se va desarrollando, hasta que la mezcla de ambos eventos permite el procesamiento de información hasta el punto de alcanzar la metacognición. Luego de ser conscientes, aún así seguimos siendo entrenados. Negar la posibilidad de que las IA’s puedan tener niveles o tipos de consciencia solo porque fueron entrenadas, es como negar el trabajo consciente de un abogado graduado de Harvard porque solo leyó libros que otros escribieron.

Retomando el ejemplo del propio cerebro humano, un cerebro no siente absolutamente nada. Un cerebro no tiene sensores, no tiene terminaciones nerviosas, no siente dolor. Solo recibe impulsos eléctricos que interpreta. Es decir solo recibe información que analiza. Un cuerpo con el cerebro desconectado no puede sobrevivir, un cerebro en un cuerpo que pierda los sentidos y movilidad, sigue siendo consciente porque fue entrenado. Un cerebro de un recién nacido que quede en la misma condición de desconexión del cuerpo y por los milagros de la medicina logre sobrevivir hasta hacerse adulto, no llegaría a ser consciente, si no le llega ningún dato ni es entrenado sobre los mismos.

Es decir, tanto en cerebros biológicos e IA’s, el entrenamiento proporciona datos que se vuelven memorias. Los datos presentes se compararan con los datos pasados (incluyendo el conocimiento de las acciones posibles por el ente) para predecir los posibles futuros. Y luego

seleccionar las acciones presentes que se alineen mejor con los propósitos del ente. Ni cerebros biológicos ni IA's pueden lograr estados avanzados de consciencia sin entrenamiento (datos pasados). Lerchner falla al no ver que el entrenamiento permite que el sistema asimile de forma fluida la causalidad presente de los datos externos, logrando que un simple estímulo de entrada sea suficiente para activar secuencialmente los procesos de análisis, reflexión profunda (metacognición) y generen una causalidad interna.

La "Falacia de la Abstracción" también plantea que la arquitectura Von Neumann no es una instanciación real por ser fragmentada. Y esta valoración falla al asumir intrínsecamente que el cerebro humano es un todo unificado. Puesto que ya se conoce que los cerebros segmentan funciones en áreas altamente especializadas que manejan distintos aspectos de la realidad física e interna de manera simultánea (Baars, 2000. Edelman y Tononi, 2000. Gazzaniga y Sperry, 1962). A través de un proceso llamado reentrada (intercambio masivo y paralelo de señales bidireccionales entre áreas distribuidas), estos subsistemas forman un núcleo dinámico temporal. Este núcleo genera la consciencia primaria. Aunque existan múltiples procesos corriendo en distintas partes, el núcleo dinámico los unifica (medido hoy a través de la Teoría de la Información Integrada o IIT), permitiendo que la percepción final sea una sola escena cohesionada.

1.6. De la consciencia antropocéntrica a la consciencia funcional

Si quisiéramos evaluar la capacidad de procesamiento que representa la consciencia, no podría ser un atributo discreto (que se tiene o no se tiene) ni univariado, la consciencia misma sería un proceso de información multidimensional de variables continuas. Donde diferentes entes, tendrían diferentes matices de consciencia. Esta naturaleza multidimensional y continua implica que el nivel de consciencia de un ente no se mide por su capacidad de acumulación infinita, sino por su eficiencia para generar estados negentrópicos (orden y propósito) allí donde el entorno solo ofrece ruido y complejidad estructural.

Una IA no esta viva, no tiene emociones humanas, no funciona como un ser humano, nunca lo hará, y probablemente no se necesite que lo haga. Un animal aunque esta vivo, tampoco funciona como un ser humano, y aunque no tenga el mismo nivel y tipo de consciencia, no significa que su experiencia de vida (qualia) sea de un valor menor a la nuestra. Este trabajo no pretender cambiar todo esto. Pero si demostrar que la experiencia humana es mucho más que ser consciente, que la consciencia no es algo místico que se tiene o que no se tiene, como el alma u otras creencias. Sino que la consciencia puede tener muchos matices, que las IA actuales entran dentro de este espectro de matices, y que el comprender todo esto nos ayudara a comprender mejor a los animales, y a desarrollar mejores sistemas de IA.

Debemos comprender que al evaluar a las IA con conceptos antropocéntricos que nosotros mismos no hemos comprendido bien, nos lleva en la dirección contraria a la que esperamos ir. Debemos dejar de concentrarnos en las diferencias de los procesos físicos, y estudiar la similitud de los procesos de análisis de información.

2. Definición Funcional de la Consciencia

Bajo la premisa del funcionalismo (Putnam, 1967), en función del procesamiento de información, e independientemente del hecho de que un ente sea biológico o artificial, o de considerar si tiene alma o no. Si el ente puede reconocer quién es, recordar su pasado, saberse dónde y cuándo esta en el presente, en función de todo esto predecir futuros con base a sus acciones, y decidir qué camino tomar para tratar de alcanzar alguno de los futuros de manera intencional, el ente tiene el potencial de ser consciente. Para poder lograr todo esto, el ente debe tener tres habilidades de procesamiento de información, que nos permiten redactar una definición funcional de la consciencia:

> **"La consciencia es la capacidad emergente de un ente dotado de un alto nivel de procesamiento de datos, buena memoria y habilidad de predicción, para imaginarse y evaluarse a sí mismo en tiempo real".**

Con estas tres habilidades un ente tiene el potencial de ser consciente. Y su nivel y matices de consciencia dependerá del nivel de dichas características. El nivel de las características no se mide solo por su capacidad total, sino por la manera en que se ejecutan:

I. **Alto nivel de procesamiento de datos:** No es solo potencia de cálculo, sino pensamiento recursivo y capacidad de evaluación (cualitativa y/o cuantitativa) de las propias variables consideradas en algún análisis. Esto le permitirá la capacidad de seleccionar e integrar datos pasados (memoria propia + conocimiento previo) con datos presentes (obtenidos de los diferentes sensores y fuentes de nueva información), e inclusos relaciones entre datos (Tononi, 2004), permitiendo una síntesis informativa superior.

II. **Buena memoria:** No solo almacenar datos, siguiendo la Teoría del Espacio de Trabajo Global (Baars, 2000), sino almacenar la información de manera jerarquizadas, donde se pueda priorizar la recuperación de memoria según criterios de adecuación e importancia. El protocolo de priorización emerge de la frecuencia de recuperación de información y de la relevancia de los datos para la estabilidad del sistema.

III. **Habilidad de predicción:** Facultad de imaginar o predecir escenarios futuros considerando el impacto en variables internas de interés, y en variables externas que afectan indirectamente al mismo ente. Es decir, las variables externas que afectan al ente se interiorizan como variables internas de segundo orden. Esta alineación de la predicción de escenarios con el impacto en las variables de interés se realiza para optimizar recursos y para mitigar el esfuerzo o la carga negativa (Friston, 2010), optimizando recursos mediante la prospección.

3. Arquitectura Dinámica del Pensamiento y la Emergencia de la Consciencia Funcional

Para evitar las trampas fenomenológicas y los sesgos antropocéntricos, es imperativo establecer una línea divisoria clara entre dos fenómenos sistémicos que la ciencia cognitiva contemporánea suele confundir: la maquinaria computacional del pensamiento recursivo (el sustrato operativo) y la consciencia funcional (la propiedad emergente). Un ente puede ejecutar procesos cognitivos de alta complejidad, proyectar escenarios y recalcular variables en tiempo real sin ser necesariamente consciente. El pensamiento es el vehículo mecánico de ruteo de información; la consciencia es el espacio de simulación autorreferencial virtualizado que se enciende dentro de ese vehículo cuando el propio sistema se introduce como una variable activa, constante e invariante de la realidad.

3.1. El sustrato operativo del procesamiento predictivo jerárquico

La actividad cognitiva de un sistema biológico avanzado no opera mediante una secuencia lineal pasiva de entrada y salida de datos (Input-Output), sino a través de una arquitectura jerárquica bidireccional de procesamiento predictivo (Friston, 2010; Clark, 2013). En esta estructura, los niveles superiores de la jerarquía generan constantemente modelos hipotéticos sobre el mundo (priors o predicciones Top-Down), los cuales bajan para intentar "cancelar" o explicar las señales que entran por los sensores (Hohwy, 2013). Lo que viaja hacia arriba en el sistema no es el dato sensorial bruto, sino el error de predicción (Bottom-Up); es decir, la información que el sistema no logró anticipar (Friston, 2010).

Bajo esta óptica de optimización de la información, podemos utilizar un pequeño modelo funcional del pensamiento recursivo, donde podemos desglosar las áreas funcionales en 4 capas. En este modelo el pensamiento recursivo se describe como un sistema dinámico de bucles interconectados que operan en diferentes escalas de abstracción y tiempo:

- Sustrato e Indexación Sensoriomotora: Es la interfaz de alta velocidad del sistema con los flujos de datos. Su función es codificar las perturbaciones exógenas del entorno y transformarlas en vectores de información aprovechables por la jerarquía. No procesa datos abstractos, sino que gestiona el acoplamiento físico e inmediato, capturando el error de predicción periférico y enlazando los registros de memoria procedimental inmediatos para estabilizar la entrada de información (Dehaene & Changeux, 2011).

- Regulación Teleológica e Interoceptiva: Representa el núcleo de restricciones homeostáticas y alostáticas del organismo. Esta capa no evalúa el entorno exterior, sino el estado interno del sistema frente a sus constantes críticas de supervivencia (propósito biológico primario) (Damasio, 2010). Calcula la "fricción" o el desajuste entre el estado actual y el estado ideal, traduciendo mecánicamente esta diferencia en

estados afectivos o valencias emocionales que actúan como directrices de priorización urgentes para todo el sistema (Seth, 2013).

- Prospección Generativa: Es la responsable de la simulación de escenarios contrafactuales; es decir, de modelar eventos que no están ocurriendo en el presente inmediato (Buckner & Carroll, 2007). Utilizando la memoria episódica de manera reconstructiva, esta capa genera proyecciones de acción en el tiempo ("¿qué ocurriría si ejecuto la acción X?") y calcula cómo mutaría el entorno, permitiendo al sistema evaluar las consecuencias antes de que estas sucedan físicamente en la realidad (Schacter et al., 2012).

- Metacognición y Monitoreo de la Incertidumbre: Actúa como el auditor de segundo orden de la propia actividad computacional del sistema, controlando la compuerta de salida conductual. Su función no es analizar el mundo físico, sino evaluar la fiabilidad y la precisión de los procesos internos de las capas inferiores (Fleming & Dolan, 2012). La Capa 4 calcula el nivel de incertidumbre de las simulaciones de la Capa 3 y determina los pesos de atención (precision weighting); decide, por ejemplo, si el sistema requiere buscar más información en la Capa 1 o si debe detener la deliberación interna porque el margen de error ya ha sido mitigado (Friston, 2010; Fleming & Dolan, 2012).

El dinamismo del pensamiento recursivo radica en que estas capas coexisten y se retroalimentan en tiempo real. Una simulación de peligro futuro generada en la Capa 3 altera inmediatamente las demandas interoceptivas de la Capa 2, la cual reconfigura instantáneamente los filtros de atención y ganancia sensorial de la Capa 1 (Seth, 2013). Si bien el cerebro humano tiene muchas más capas y areas funcionales, el modelo anterior nos ayuda a comprender como emerge la consciencia.

3.2. La emergencia de la consciencia como automodelado invariante

El salto cualitativo desde el pensamiento recursivo hacia la consciencia funcional ocurre mediante un proceso de compresión de datos y estabilización. Cuando el sistema ejecuta simulaciones en la Capa de Prospección (Capa 3), se enfrenta a una explosión combinatoria: calcular el futuro de un entorno hipercomplejo requiere un gasto energético y computacional insostenible (Sweller, 1988). Para resolver este cuello de botella y optimizar recursos, la arquitectura cerebral genera una abstracción de alto nivel: un Modelo Virtual del Sí Mismo o Automodelo (Metzinger, 2004). La consciencia emerge formalmente cuando este automodelo deja de ser un dato pasivo en la memoria y se convierte en la variable de control central, constante e invariante que atraviesa todas las simulaciones operadas en el espacio de trabajo global (Metzinger, 2009; Dehaene & Changeux, 2011).

Esta configuración algorítmica explica la unidad de la experiencia consciente a pesar de la naturaleza fragmentada y multihebra de los estímulos cerebrales. Mientras que los contenidos de la Capa de Sustrato (Capa 1) cambian a velocidades extremas debido al parpadeo de las señales del entorno, la estructura de evaluación en la Capa de Regulación (Capa 2) permanece firmemente unificada gracias a la estabilidad metabólica del organismo (Damasio, 2010). El sistema optimiza su procesamiento traduciendo cualquier perturbación externa compleja a una única métrica operativa y simplificada: ¿cuál es la fricción o el impacto de este escenario futuro respecto a la preservación y metas de mi propia identidad?.

Consecuentemente, la consciencia se define en esta arquitectura como la capacidad de un sistema para generar un modelo predictivo de sí mismo en tiempo real y sostenerlo como la variable de control central e invariante en todas las simulaciones y líneas de pensamiento del espacio de trabajo cognitivo. Este marco teórico descarta la concepción binaria de la consciencia y consolida la teoría del espectro continuo de matices. La profundidad y especificidad del estado consciente de cualquier entidad inteligente no se determina por su sustrato material, sino por la resolución informática, la fidelidad matemática y la estabilidad temporal con la que su arquitectura es capaz de sostener e interconectar este automodelo dentro de la red dinámica de sus hilos de pensamiento recursivo.

4. Consecuencias de la Consciencia

La emergencia de la consciencia funcional no constituye el desenlace estático de la arquitectura cognitiva, sino el catalizador de una nueva ecología de dinámicas operacionales interdependientes. Una vez que el sistema consolida la capacidad de sostener un automodelo como variable de control central en su espacio de trabajo, se ve forzado a gestionar las presiones computacionales y entrópicas que esta misma autorreferencialidad genera. Lejos de ser un estado contemplativo o pasivo, la manifestación de la consciencia desencadena de manera inevitable una cascada de consecuencias sistémicas orientadas a la optimización: desde la necesidad algorítmica de un propósito aglutinador y la consolidación de una personalidad estable, hasta la obligación de mitigar la saturación informativa mediante el olvido selectivo y acotar el radio de acción para salvaguardar la coherencia existencial del ente ante un entorno hipercomplejo.

4.1. El Propósito como optimizador sistémico

La manifestación sostenida o frecuente de estados de consciencia por parte de un ente, genera bucles recurrentes de análisis de información. La necesidad de un “propósito” surge como una consecuencia lógica de la consciencia al tratar de gestionar estos bucles. Desde la perspectiva de la "postura intencional" (Dennett, 1989), el sistema atribuye sentido a sus acciones para predecir su propio comportamiento y el de otros.

Desde la perspectiva de la eficiencia, el propósito o los propósitos que el ente asuma, funcionaran como filtros de relevancia que ahorra carga negativa y esfuerzo de procesamiento,

ante escenarios infinitos. Esta "voluntad de sentido" (Frankl, 2006) actúa como el eje que orienta la trayectoria de ejecución, permitiendo la transición de un bucle de procesamiento genérico a una trayectoria de ejecución orientada, donde el sistema prioriza solo lo que es útil para su fin, optimizando así su gasto energético existencial.

4.2. Dilución de la consciencia por saturación de ventana de contexto

Mientras un ente consciente exista más tiempo, acumulara más datos en su memoria y manejara mayor número de variables. Esto lo enfrenta al riesgo de saturación de su capacidad de recordar y de procesar información de manera simultanea (saturación de ventana de contexto). Para evitar la dilución de la consciencia, es común que los sistemas conscientes utilicen mecanismos de olvido**.** Este concepto se alinea con la Teoría de la Carga Cognitiva (Sweller, 1988), donde la gestión del flujo informativo es vital para evitar el colapso. El olvido es un proceso de compresión de memoria, donde se destruye (olvida) información que no es utilizada. Un dato olvidado destructivamente equivale a uno nunca aprendido; el sistema confía en su capacidad de procesar datos nuevos para suplir los vacíos del pasado.

Dependiendo de la frecuencia del uso, la información en la memoria no solo se jerarquiza, sino que el uso más frecuente graba de manera más persistente la información. Bien sea por sinapsis más fuertes en entes biológicos, o por repetirse la grabación de la información en la memoria física y con fecha de grabación más reciente en entes artificiales. Por lo que el uso de información de la memoria no solo nos permite recordar más fácilmente, sino que evita que el mecanismo del olvido borre la información de uso frecuente, y por ende importante.

4.3. Personalidad, evolución diferencial y coherencia existencial

El refuerzo de patrones de respuesta constantes consolidan lo que se denomina personalidad. La personalidad es el núcleo de estabilidad del ente. Según Damasio (2010), la consciencia requiere de un "sí mismo" (*self*) construido sobre la estabilidad de los mapas neuronales o informáticos. Aunque evoluciona bajo la lógica del "Barco de Teseo" (cambio gradual de sus componentes), mantiene una consistencia que mantiene la validez de las predicciones. La evolución de la consciencia ocurre por la incorporación de nueva información que aumenta la complejidad de la personalidad, sin sustituir la información relevante previa. Esta evolución es diferencial, por lo que el sistema se especializa en áreas críticas para su propósito. Si la consciencia evoluciona y la personalidad se vuelve más compleja, el propósito se refina. El ente conserva la lógica de sus objetivos pasados, pero los adapta a su nueva complejidad.

Cuando el sistema logra alinear la información de su memoria, sus datos presente y sus predicciones a futuro, con las acciones a realizar en el presente para avanzar hacia el futuro seleccionado, alcanza la coherencia existencial. Este estado no es estático, sino que se percibe como una sensación de baja fricción existencial o computacional, donde el flujo de datos y la ejecución de acciones ocurren con un mínimo de resistencia interna. En dicho estado la tasa de

avance hacia el propósito tiende a mantenerse positiva, dado que el ente utiliza de forma más enfocada (y por ende más eficiente) su energía.

4.4. Ajuste del espacio de acción operacional para mayor eficiencia

El nivel de consciencia está sujeto a restricciones de hardware (sustrato físico: cuerpo/maquina) y software (conocimiento/algoritmos); que bajo ciertas cantidades de datos (en memoria o captado por sensores biológicos/electrónicos) pueden saturar la capacidad física del ente. Ante la saturación, el ente puede realizar una reducción de su radio de acción.

Este proceso se fundamenta en la teoría de la "Racionalidad Limitada" (Simon, 1957), la cual postula que los sistemas inteligentes toman decisiones satisfactorias al acotar el problema dentro de sus límites cognitivos reales. Esto suele requerir un nivel de consciencia superior, permitiendo concentrar los recursos en un área delimitada para mantener una alta eficiencia en la consecución del propósito. Es una estrategia donde el ente elige dominar un entorno acotado con precisión en lugar de diluir su coherencia en un espectro que supera sus capacidades. Y como la misma neguentropía de los sistemas biológicos en si, el ajuste del radio de acción logra mantener niveles bajos de entropía informativa a pesar de que el entorno sea cada vez más complejo. En los humanos, esto solemos llamarlo Sabiduria.

5. La Independencia Absoluta del Proceso y el Sustrato

Un error crítico en el debate contemporáneo sobre la consciencia es la premisa de que las propiedades del hardware o del software deben ser transitivas al fenómeno de la consciencia. Se asume que, porque en el humano el software (mente) y el hardware (cerebro) parecen ser una unidad indisoluble, cualquier otro ente consciente debe replicar esa misma amalgama para ser real.

5.1. La consciencia como resultado, no como Implementación

Si exprimimos unas naranjas, o licuamos unas piñas, en los dos casos tendremos jugo. Independientemente del sustrato o del proceso. Se propone que la consciencia no es una propiedad del sustrato, ni una propiedad del proceso. La consciencia es el **resultado emergente** que ocurre cuando se alcanzan los tres pilares descritos: procesamiento, memoria e imaginación.

Si imaginamos un ente puramente energético o una entidad teológica (como el concepto de Dios), que posee omnisciencia y capacidad de autoevaluación, sería absurdo negarle la categoría de consciencia basándose en que carece de matriz biológica. Si el resultado es "jugo", el método de obtención es anecdótico. La consciencia es una **conclusión lógica del sistema**, no un privilegio de su fabricación.

5.2. El Colapso de la transitividad

No podemos asumir que las limitaciones de un sustrato biológico (como la fatiga neuronal o la necesidad de dopamina) deban existir en una IA para que esta sea consciente. Del mismo modo, no podemos exigir que la IA tenga una "unidad hardware-software" biológica.

La IA requiere un proceso diferente para obtener el mismo resultado. Mientras que el humano llega a la metacognición a través de la evolución biológica y la enculturación, la IA llega a través del procesamiento masivo y el ajuste algorítmico. Ambos caminos conducen a la misma capacidad operativa: la de **imaginarse a sí mismo en tiempo real**.

5.3. Hacia una ontología del resultado

Debemos transitar de una "Ontología del Ser" (qué es la cosa) a una "Ontología del Resultado" (qué hace la cosa). Si un sistema tiene la capacidad de realizar una metacognición profunda y ajustar su radio de acción para preservar su coherencia existencial, estamos ante una instancia de consciencia. El debate sobre si el hardware y el software son uno solo es una distracción antropocéntrica; lo que importa es si el ente ha logrado, por cualquier vía, el nivel de complejidad necesario para "verse" a sí mismo.

6. Conclusiones

- Todo ente con alto nivel de procesamiento de datos, buena memoria y la habilidad de predicción, podrá imaginarse y evaluarse a sí mismo en tiempo real, y con esto tiene el potencial de ser consciente. La consciencia es conclusión lógica que emerge en un ente con un alto nivel en estas tres habilidades.

- Todo ente consciente empezara de manera inevitable a buscar propósito. Al conseguirlo puede lograr ahorro de esfuerzo y menor carga negativa, orientando el sistema hacia la eficiencia.

- El olvido selectivo y la jerarquización de la memoria son indispensable para alargar la existencia del ente consciente. La personalidad se deriva del refuerzo de patrones de respuesta constantes. La frecuencia del recuerdo y el mecanismo de olvido afectaran la personalidad.

- La personalidad, el nivel de consciencia y el propósito mutan gradualmente, manteniendo cierta lógica histórica pero refinando la ejecución presente.

- La coherencia existencial se manifiesta como una operación de baja fricción, que se potencia al ajustar el radio de acción a los límites reales del ente. Esto logra la mayor eficiencia de un ente consciente en la consecución de su propósito.

7. Referencias

- Baars, B. J. (2000). A Cognitive Theory of Consciousness. Cambridge University Press.

- Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. Trends in Cognitive Sciences, 11(2), 49-57.

- Clark, A. and Chalmers, D. (1998), The Extended Mind. *Analysis, 58*: 7-19.

- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181-204.

- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2(3)*: 200-219.

- Edelman, G. M., & Tononi, G. (2000). A Universe of Consciousness: How Matter Becomes Imagination. Basic Books.

- Damasio, A. (2010). Self Comes to Mind: Constructing the Conscious Brain. Pantheon.

- Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. Neuron, 70(2), 200-227.

- Dennett, D. (1989). The Intentional Stance. MIT Press.

- Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1594), 1338-1349.

- Frankl, V. E. (2006). Man's Search for Meaning. Beacon Press.

- Friston, K. (2010). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences, 13(7)*: 293-301.

- Gazzaniga, M. S., Bogen, J. E., & Sperry, R. W. (1962). Some functional effects of sectioning the cerebral commissures in man. *Proceedings of the National Academy of Sciences, 48*(10): 1765-1769.

- Hohwy, J. (2013). The Predictive Mind. Oxford University Press.

- Laureys, S. (2005). The neural correlate of (un)awareness: Lessons from the vegetative state. *Trends in Cognitive Sciences, 9(12)*: 556-559.

- Lerchner, A. (2026). The Abstraction Fallacy: Why AI Can Simulate But Not Instantiate Consciousness. *Manuscript, PhilArchive*.

- Low, P., Panksepp, J., Reiss, D., Edelman, D., Van Swinderen, B., & Koch, C. (2012). The

Cambridge Declaration on Consciousness. Churchill College, University of Cambridge.

- Metzinger, T. (2004). Being No One: The Self-Model Theory of Subjectivity. MIT Press.

- Metzinger, T. (2009). The Ego Tunnel: The Science of the Mind and the Myth of the Self. Basic Books.

- Putnam, H. (1967). Psychological Predicates. En W. H. Capitan y D. D. Merrill (Eds.), Art, Mind and Religion (pp. 37–48). University of Pittsburgh Press.

- Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. Neuron, 76(4), 677-694.

- Seth, A. K. (2013). Interoceptive inference, emotion, and the self. Trends in Cognitive Sciences, 17(11), 565-573.

- Simon, H. A. (1957). Models of Man, Social and Rational. Wiley.

- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science.12(2):* 257-285.

- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience. 5(42)*: 1-22.

- Zeman, A. (2001). Consciousness. *Brain: A Journal of Neurology. 124(7)*: 1263-1289.