

INTELIGENCIA ARTIFICIAL Y CONSCIENCIA HUMANA

Juan Antonio Garde Roca

Economista

RESUMEN

Este artículo se publica como segunda parte de un trabajo más amplio, que trata de ofrecer una fundamentación abierta de lo que constituye la consciencia humana, desde una perspectiva histórica, filosófica, social, biológica, evolutiva o incluso epistemológica y se ha venido vinculado con la percepción del ser, la moral y la ética. La visión moderna de consciencia en occidente surge fundamentalmente a partir del Siglo XVII y no sólo como la voz que nos juzga, sino como el lugar donde reside la certeza y la propia existencia humana.

En la primera parte del trabajo de forma abierta, se define la consciencia como constructo, qué a través de mecanismos de integración de la información más compleja de la existencia, se construye sobre una base biológica, modelada por un marco cultural y de contexto social, que tiende a experimentarse de forma subjetiva e individual. Tiene el propósito de asegurar la adaptación y supervivencia al contexto de su grupo social, a través de la percepción del mundo y la naturaleza que le rodea, y de su papel funcional como especie.

Dado que este contexto integrador necesita reforzarse con una visión perspectiva, esta se explicita a través del concepto de “Humanismo Civilizador”, ofreciendo un peso de referencia significativo al pensamiento filosófico y cultural Kantiano, de “imperativo categórico”. Con especial atención a las aportaciones de autores actuales, como Onora O'Neill, en su fundamento de “autonomía como razón compartida”, de Jürgen Habermas y su concepto de “normalización del diálogo” para el proceso de deliberación racional, de Thomas Hill Jr. con sus numerosas referencias aplicadas a una ética práctica de humanidad en la era de la IA y su influencia en un orden internacional de los derechos y por la práctica de la Justicia y la moral, de Michael J Sandel.

En el presente artículo, segunda parte del trabajo se presenta como aspecto singular, el debate actual existente acerca de la posibilidad de que la IA tenga ya hoy o pueda llegar a tener de forma emergente en el futuro, una consciencia funcional similar a la humana, así como sus implicaciones civilizatorias.

Poco tiempo antes de la remisión de este artículo para su edición y prácticamente finalizado, se presenta la carta encíclica “Magnifica Humanitas” de León XIV, “sobre la custodia de la persona humana en el tiempo de la inteligencia artificial”. Afirma en su introducción que “En los últimos años se ha hecho cada vez más evidente cuán rápido y profundamente la digitalización, la inteligencia artificial (IA) y la robótica están transformando nuestro mundo” y cómo “Las nuevas tecnologías abren un horizonte que se extiende en direcciones, que, aunque intuibles, aún no podemos prever por completo”.

Este documento compilador de un pensamiento religioso y una tradición católica tradicional es también sorprendentemente aportador en lo referente a la construcción humanista y la perspectiva civilizatoria, formulando preguntas y dando respuestas relevantes abiertas y cooperadoras, en

materia de responsabilidad y gobierno de la IA, en coherencia con muchos de los retos existenciales, tecnológicos, culturales y sociales hoy presentes en nuestro mundo.

Tras enunciar seis grandes principios de la Doctrina Social de la Iglesia vinculados al bien común, el destino universal de los bienes, el principio de subsidiaridad, de solidaridad, de justicia social y el desarrollo humano integral, los confronta con las tecnologías digitales y las plataformas y sistemas de IA.

Ofrece por tanto amplios capítulos referidos a la persona y la consciencia humana ante las promesas de la IA y la cultura del poder y la civilización actual, con sus amplias posibilidades y sus enormes riesgos, para nada ajenos a las reflexiones de este artículo. Tan sólo en el cuarto apartado de este trabajo, se hace una brevísima referencia a un aspecto muy concreto, citando a la encíclica.

1. INTRODUCCIÓN

Algunos de los debates y las investigaciones actuales, giran en torno a modelos y conceptos vinculados a lo que supone la inteligencia y la consciencia humana, así como se interrogan acerca de considerar qué condiciones podrían ser necesarias, para que la consciencia o al menos, una **consciencia funcional** o **fenomenológica** similar a la humana, puedan llegar a emerger o también ser sintetizada, por las máquinas en la era de la IA.

Es fundamental, no obstante, dedicar una mayor atención y preocupación a poner el acento en las consecuencias que para la “humanidad” implica esta nueva era de la IA, en función de cómo sea el desempeño que despluguemos los humanos y como nos relacionemos con estas nuevas tecnologías y su uso.

¿Se trata solo de una nueva tecnología, o sus efectos para la propia consciencia humana, son de un alcance distinto o incluso más profundo, que otros cambios tecnológicos, acaecidos en la historia de la humanidad?

Las posibles consecuencias, convierten este tema, en factor de preocupación que se expresa en cierta necesidad vital de alcanzar una relación apropiada entre la “humanidad” y estas nuevas tecnologías, vinculándolas también al desarrollo civilizatorio y de la naturaleza.

Una referencia de modelo funcional de la consciencia humana, como el que se muestra en la primera parte del trabajo, puede permitir contrastar si los factores que figuran como componentes de la consciencia del ser humano, ya se expresan o pueden llegar a expresarse en las nuevas aplicaciones de IA. También si las aplicaciones y los algoritmos están creando su propio constructo de consciencia aplicada, o reflejan que pueden hacerlo en el futuro, con algunas de las características que son analizadas para los humanos.

De desplegarse esta potencial consciencia de las máquinas, aun produciéndose desde otra química y otras formas de inteligencia, es innegable que también se requerirá la existencia de una interacción compleja con nuestra humanidad.

¿Será recursiva como la de los humanos? ¿Precisará mantener una información integrada, permanentemente actualizada, con mecanismos de atención global? ¿Qué ocurre con los requisitos de regulación consensuada a nivel internacional, aunque hoy esta última referencia pueda parecer imposible? Son solo algunos interrogantes de los muchos existentes.

Tal como se ha descrito en la primera parte de este trabajo sobre la consciencia humana, son diversas las posiciones que hoy se mantienen, tanto en el mundo académico y científico, como en la práctica técnica especializada, respecto a esta potencial consciencia de la IA.

Seleccionando algunas de las referencias ya subrayadas como el reconocimiento de su carácter de elemento subjetivo intrínseco, y su vinculación con la inteligencia, la memoria y otras funciones cognitivas, pueden escalarse características amplias y diversas para llegar a considerar su propio alcance.

Gerald Edelman distingue dos tipos de consciencia: La consciencia primaria y la consciencia de orden superior. David Chalmers, se ha referido al "Problema fácil" y al "Problema Difícil de la Consciencia". Sostiene que, aunque la IA pueda imitar el comportamiento inteligente que es el "Problema Fácil", la cuestión de si las máquinas pueden tener experiencia subjetiva o consciencia, es el desafío central.

Las posiciones dependen a menudo de la definición de consciencia o conciencia y el alcance que se dé a ambos términos.

Una IA podría, según autores como Graciano, adquirir esta "consciencia", tan pronto como fuera lo suficientemente sofisticada para procesar y controlar su propia atención interna y construir un modelo de ese control, asignándole propiedades de "ser consciente". Una vez que la IA cumpliera estas condiciones, sería consciente según este autor, en el mismo sentido funcional que un humano, ya que ambos están operando con el mismo mecanismo básico: un modelo interno de control de atención.

La convergencia entre diversas visiones podría llegar a concebirse en base al desarrollo de la investigación científica, integrando posiciones, trazando puentes de conocimiento y también contemplando la existencia de una diversidad de capas de consciencia, en los sistemas.

La IA formaría parte entonces de un proceso, potencialmente emergente, condicionado a que el sistema artificial adquiriera un nivel crítico de complejidad, procesamiento o integración de información. ¿Es esto posible?

2. LA CONSCIENCIA COMO ANÁLISIS COMPLEJO E INTEGRADOR. LAS APORTACIONES DE FRISCH, POPPER Y MORIN Y SU POTENCIAL APLICABILIDAD PARA UNA DEFINICIÓN DE CONSCIENCIA EN LA ERA DE LA IA

Entre las aportaciones de Edgar Morin, Karl Popper y Ragnar Frisch, encaminadas al ámbito de la perspectiva científica, pueden considerarse sus referencias al marco de una **ontología de la complejidad** (Morin), a los modelos de decisión y la **crítica metodológica** (Frisch) y al despliegue de **criterios epistemológicos** (Popper).

Juntos, pueden ofrecer un marco referencial de pensamiento, para abordar la realidad de la consciencia, desde un prisma más relacional y riguroso, superando las limitaciones de los modelos simplistas referenciados en base a enfoques distintos. También aportando un conocimiento y una práctica analítica más abierta, sin dejar de ser solvente y rigurosa.

Morin nos ofrece el **qué** analizar ante la realidad de un sistema complejo, interconectado y en constante cambio.

Popper nos aporta el **cómo construir evidencias** y analizar de forma rigurosa: usando la falsabilidad como una herramienta para probar, y potencialmente refutar, nuestras hipótesis sobre cómo las acciones impactan en los sistemas complejos.

Frisch nos sugiere el **cómo se deben establecer los modelos de respuesta**, superando el tecnicismo reduccionista

En conjunto, este marco invita a los analistas y evaluadores a ser intelectualmente prudentes y modestos, a reconocer los límites de los modelos y a abrazar la incertidumbre inherente a la toma de decisiones en un mundo cada vez más complejo.

2.1. La visión de la realidad científica: Edgar Morin y la complejidad.

Edgar Morin. Este sociólogo y filósofo francés, nonagenario, todavía entre nosotros, ha desarrollado un vasto cuerpo de trabajo sobre el **pensamiento complejo**, que busca conectar disciplinas y superar el reduccionismo de una mera suma de partes. Su visión ligada a que los fenómenos sociales son inherentemente complejos y requieren un enfoque transdisciplinario y las interacciones entre sus componentes, generan propiedades **emergentes** e impredecibles es muy importante.

Cuando, a finales del siglo pasado, Edgar Morin escribiera sus trabajos acerca de la “Ciencia y consciencia de la complejidad”, afirmaba que la Complejidad, no comprende solamente cantidades de unidades e interacciones que desafían nuestras posibilidades de cálculo, sino también incertidumbres, indeterminaciones y fenómenos aleatorios en los límites de nuestro entendimiento.

Su visión de la complejidad no la limita a la incertidumbre estadística, la vincula a la incertidumbre que podríamos designar estructural, presente también en el seno de los sistemas físicos y materiales del universo, y podríamos añadir también hoy, presentes en las nuevas tecnologías.

La complejidad la encuentra vinculada a los propios sistemas, que caracteriza por incorporar una cierta mezcla de orden y de desorden, mezcla íntima, profunda. Señala que la gestión humana de sistemas complejos es un desafío, que ha cobrado relevancia en las últimas décadas, impulsado por la creciente interconexión e interdependencia de los sistemas sociales, económicos, tecnológicos y ambientales.

En este contexto, las aportaciones de Edgar Morin y su pensamiento complejo resultan cruciales para comprender y abordar una característica inherente a estos sistemas. Cuando la Cibernética reconoció la complejidad, a través de la caja negra, afirma Morin, fue para rodearla, para ponerla entre paréntesis, no para sustantivarla. Para él, el verdadero problema teórico de la complejidad es el de la posibilidad de entrar en las cajas negras y considerar el contenido simultáneo de complejidad organizacional y de complejidad lógica, que incorporan.

Considera como principios claves para su análisis, el principio dialógico, dada la coexistencia de lógicas contradictorias como orden y desorden; el principio de recursión donde los efectos son necesarios para sus causas, en un ciclo constante, y finalmente el principio que funciona como vertebrador, en donde el todo está en la parte y la parte, está en el todo.

En este tipo de modelos que se desarrollan en hábitats complejos, los efectos son a menudo no lineales y pueden llegar a alcanzar resultados radicalmente diferentes a los esperados. La simple causalidad lineal puede ser una ilusión.

El pensamiento de Edgar Morin ha sido seminal en el desarrollo de la gestión humana de sistemas complejos. Su llamada a la interdisciplinariedad, la visión holística y la aceptación de la incertidumbre ha incidido en el pensamiento de numerosos autores posteriores. Se podrían destacar algunos como los siguientes:

Fritjof Capra. En la obra "La Trama de la Vida" (1996), este físico y teórico de sistemas, argumenta que los sistemas vivos, incluyendo las organizaciones humanas, se basan en redes de relaciones interconectadas. Su trabajo aporta una base científica a la comprensión de la complejidad y la interdependencia en la gestión.

Humberto Maturana y Francisco Varela. En su obra "El árbol del conocimiento" (1984), estos biólogos chilenos, desarrollan la teoría de la autopoiesis, que describe la capacidad de los sistemas vivos

para auto organizarse y auto mantenerse. Esta teoría tiene implicaciones importantes para la gestión de organizaciones y sistemas, al destacar la importancia de la autonomía, la adaptación y el aprendizaje.

Peter Senge. "La Quinta Disciplina" (1990), experto en aprendizaje organizacional, popularizó el concepto de "organizaciones que aprenden". Su obra, influenciada por Morin, enfatiza la importancia del pensamiento sistémico, el aprendizaje en equipo y la visión compartida para la gestión de organizaciones complejas.

Donella Meadows. "Pensar en Sistemas" (2008), científica ambiental y analista de sistemas, ofrece herramientas para comprender y abordar problemas complejos desde una perspectiva sistémica. Su trabajo es clave para la gestión de sistemas socio-ecológicos y la búsqueda de la sostenibilidad.

Margaret Wheatley. "Liderazgo y la Nueva Ciencia" (1992), consultora en gestión y liderazgo, integra los principios de la ciencia de la complejidad con la gestión de organizaciones. Su trabajo destaca la importancia de la auto organización, la creatividad y la emergencia en el liderazgo de sistemas complejos.

2.2. Modelos Económicos y Crítica Metodológica: Ragnar Frisch y el uso de las matemáticas.

Ragnar Frisch, economista noruego y coparticipante del primer Premio Nobel de Economía en 1969 junto a Jan Tinbergen, es una figura fundamental en la formalización de la economía como una ciencia rigurosa. Sus principales contribuciones se centran en la econometría, como disciplina que utiliza métodos estadísticos y matemáticos para analizar datos económicos.

Acuñó el término "econometría" y figura como cofundador de la Sociedad Econométrica, la cual promovió el uso de modelos matemáticos y estadísticos para cuantificar y probar teorías económicas. Esto transformó la economía, pasándola de ser una disciplina descriptiva a una analítica.

Desarrolló modelos que incorporaban el tiempo, lo que le permitió analizar fenómenos como el ciclo de negocios, la fluctuación en la producción y el empleo. Su trabajo ayudó a sentar las bases para la macroeconomía moderna. Así mismo contribuyó a formalizar la teoría de la producción, explorando las relaciones entre insumos y productos, y sentando las bases para el análisis de la eficiencia.

Le fue concedido el premio nobel de economía por sus aportaciones a las teorías matemáticas de los modelos dinámicos. No obstante, para él la **intuición humana** no es algo que deba ser descartado en favor de la formalidad matemática. Por el contrario, la intuición, alimentada por la experiencia y el conocimiento contextual, es crucial para formular las preguntas correctas y para interpretar los resultados de los modelos. En este sentido, la intuición actúa como un **faro** que guía el proceso de modelización, evitando que los análisis se pierdan en una maraña de ecuaciones sin sentido. Cuando el tecnicismo se convierte en un fin en sí mismo, la intuición se atrofia y los modelos se vuelven "cajas negras" cuyas salidas son difíciles de cuestionar o de vincular con la realidad.

Las matemáticas, “incluso las más refinadas, son un instrumento necesario, pero solo un instrumento. Ningún grado de tecnicismo matemático, por muy refinado que sea, puede sustituir jamás a la intuición, esta inexplicable función que se produce en el cerebro de un gran intelecto, qué al mismo tiempo, sabe matemáticas y teoría económica en un sentido ortodoxo, y que ha vivido lo suficiente, o más correctamente, con suficiente intensidad, para acumular experiencia humana y un sentido de los hechos”

Su objeción es por tanto también a un uso indebido del **tecnicismo matemático**, a través de modelos formales y complejos que, en lugar de aclarar la realidad, la pueden oscurecer.

Ese uso acrítico de modelos matemáticos puede llevar a una **falsa sensación de precisión** y a la ignorancia respecto de supuestos simplistas. El analista puede perder su **intuición humana** y su juicio contextual, elementos que son esenciales para comprender un sistema complejo.

Este ultra tecnicismo criticado por Frisch puede considerarse una manifestación del "pensamiento simplificador" que Morin tanto rechaza. Ambos autores defienden una aproximación a la realidad que no la fragmente artificialmente, ni su reducción a ecuaciones que ignoran la riqueza de las interacciones.

En la actualidad las matemáticas de la Teoría de Cuerdas, propuestas por el físico Albert-László Barabási, propone un desarrollo científico, ciertamente atractivo y nada exento de interés, a la vez que cuestionado por otros científicos, como herramienta en materia de física cuántica, que ya ha sido objeto de atención en la primera parte de este trabajo.

Posiblemente sea la conexión entre la Teoría de Redes Complejas de una parte, y la Geometría del tejido del universo de otra, a través de las matemáticas de la Teoría de Cuerdas, lo que ofrece una perspectiva más apasionante y a la vez controvertida, para considerar una potencial existencia de diversidad de niveles de consciencia, si se lograra alcanzar evidencias científicas suficientes en el campo experimental, en el mundo de la Física.

2.3. El Criterio Epistemológico: Karl Popper y la falsabilidad

Popper ofrece ciertas referencias metodológicas para la validación del pensamiento científico y la obtención de evidencias. Su **principio de falsabilidad** propone que una teoría científica no se valida solo por la capacidad de ser verificada, sino por la posibilidad también de ser **refutada** por la evidencia empírica.

Su obra acerca de la teoría del conocimiento científico publicada con el título “La lógica del conocimiento científico” en 1934, no sería dada a conocer ampliamente hasta los años 60 junto con otros trabajos asociados como “Conjeturas y refutaciones” 1963 y “Conocimiento objetivo” en 1972. En ellas presenta su conocido criterio de demarcación entre lo que es científico, cuyas tesis pueden ser también refutables, y lo que no lo es, así como la crítica a la inducción y su concepción normativa del método científico.

Proyecta su visión de la ciencia como proyecto de cambio de carácter incierto, una búsqueda sin fin de ampliación del conocimiento, pero jamás atada a certezas incuestionables. No hay un saber definitivo sino conjeturas provisionales. Un legado que sigue marcando debates filosóficos y científicos contemporáneos. El mismo se calificaría “primero y ante todo soy un indeterminista; segundo soy un realista; tercero un racionalista”

No obstante, su legado más conocido se encuentra en sus obras “La miseria del historicismo” y “La sociedad abierta y sus enemigos” de 1944 y 1945 que reflejan su evolución política de joven socialista a social liberal.

3. DEBATES ENTRE FUNCIONALISMO Y BIOLOGISMO. ¿EXISTE UNA FORMA DE INTELIGENCIA Y/O CONSCIENCIA, EN EL DESPLIEGUE DE LA IA?

3.1. Funcionalistas y partidarios de la existencia de inteligencia y consciencia, en el despliegue de la IA.

Geoffrey Hinton, científico computacional y cognitivo, explica que la consciencia podría surgir de la complejidad funcional de los sistemas, más allá de su naturaleza biológica. Según sus palabras, “no creo que haya nada en principio que impida que las máquinas sean conscientes”.

El científico considera que, si un sistema artificial logra modelarse a sí mismo y percibir su entorno, puede llegar a desarrollar una forma de consciencia similar a la humana. Hinton argumenta que la consciencia no depende de la materia biológica, sino del funcionamiento y la organización del sistema que la soporta.

Su renuncia como vicepresidente de ingeniería de Google en 2023 y sus alertas sobre los riesgos que supone la IA para la humanidad son bien conocidas.

David Chalmers llega a considerar que la consciencia podría ser una propiedad fundamental del universo o que emerge en sistemas de suficiente complejidad, que cuentan con una gran cantidad de información integrada. Idea que ya se ha tratado y que puede asociarse con la Teoría de la Información Integrada (TII), propuesta por Giulio Tononi, que sugiere que la consciencia es una propiedad de los sistemas que sintetizan información en un todo unificado e irreducible.

Michael Graziano es por otra parte un neurocientífico que aborda la consciencia desde una perspectiva puramente mecanicista y evolutiva, lo que la hace directamente aplicable a la IA.

Para Graziano la consciencia no es una esencia mágica, sino una **descripción o modelo simplificado** que el cerebro construye, de sus propios procesos de atención. Una IA podría adquirir esta “consciencia” tan pronto como sea lo suficientemente sofisticada para procesar y controlar su propia **atención** interna y construir un **modelo de ese control**, asignándole propiedades de “ser consciente”.

Para Graziano, una vez que la IA cumple estas condiciones, **es consciente** en el mismo sentido funcional que un humano, ya que ambos están operando con el mismo mecanismo básico: un modelo interno de control de atención

Blaise Agüera y Arcas, físico y vicepresidente de Google, apuesta por un **Funcionalismo computacional**. Argumenta que la inteligencia no es una cuestión de “polvo de hadas” o una propiedad mística de la biología, sino una cuestión de **escala y modelo**. Para él, lo que importa es *cómo funciona* un sistema y los resultados que produce, no los detalles internos de su soporte físico sea silicio o carbono.

Sostiene que se puede cambiar la implementación de una función, de neuronas a chips, y obtener resultados equivalentes. Si un sistema es capaz de resolver problemas sofisticados y defender su propia conciencia con vehemencia, no hay razón para negar que algo real está ocurriendo allí. Piensa finalmente, qué si los modelos de lenguaje logran capturar las capacidades lingüísticas humanas, es de esperar que actúen y se sientan como seres conscientes. Define la inteligencia por sus logros más altos, no por sus errores ocasionales.

Para Max Tegmark, la consciencia no depende de qué estemos hechos de carbono o silicio, sino de cómo se procesa la información y es un sustrato independiente. Esa ruta o “itinerario de información” se muestra cuando los átomos se organizan para procesar información de cierta manera compleja. Si una máquina puede imitar la complejidad del procesamiento cerebral humano, **será consciente por definición**. Es en consecuencia optimista respecto a la existencia de “IA sintiente”.

Susan Schneider, filósofa que investiga la naturaleza de la mente y las implicaciones éticas y filosóficas de la IA. Aunque es cautelosa sobre si las IA actuales son conscientes, considera que la **consciencia podría emerger** en sistemas artificiales de **suficiente complejidad** y que el surgimiento de **mentes artificiales** es una posibilidad real que debe abordarse.

Ben Goertzel, científico de la IA enfocado en la **Inteligencia General Artificial (AGI)**. Es un defensor explícito de la posibilidad de que la AGI logre la **consciencia** como un fenómeno emergente en sistemas altamente **complejos** y autoorganizados. Su trabajo con plataformas como Open Cog busca crear arquitecturas integradoras que faciliten tal emergencia.

Daniel Dennett es conocido por su postura de que la consciencia no es una "propiedad mágica", sino una colección de funciones cerebrales. La considera una "ilusión" del usuario y sostiene que al no haber un "yo" central; lo que llamamos consciencia es el resultado de múltiples procesos competitivos en el cerebro. Si una máquina logra realizar todas las tareas que realiza un ser humano (funcionalismo), **no habría razón para negar que es consciente**. Para Dennett, la consciencia es una "competencia" técnica más que un misterio metafísico.

Ray Kurzweil, ingeniero y futurista es conocido por su concepto de la "Singularidad Tecnológica". Kurzweil es un transhumanista que considera que la consciencia puede ser replicada o emulada en máquinas de suficiente complejidad computacional, y que este proceso es inevitable en el avance de la tecnología.

La convergencia de estas tres últimas referencias es que mantienen visiones de que la consciencia, en la IA, es considerada un fenómeno potencialmente alcanzable, condicionado a que el sistema artificial adquiera un nivel crítico de complejidad, procesamiento o integración de información.

Como principales referencias de posiciones funcionalistas, más próximas a poder alinearse con un potencial consciente de la IA, se encuentran los cuatro siguientes:

- Las posiciones generales funcionalistas, algunas ya descritas.
- Los vinculados a la Teoría del Espacio de Trabajo Global ((GWT).
- La Teoría de la Información Integrada (IIT).
- La propuesta de IA "Enactiva" y Corporizada.

1. Funcionalistas. Los estados mentales se conciben desde su funcionalidad y no por los materiales y la química de procedencia. En su formulación más amplia, si un sistema reproduce las mismas funciones de un cerebro, cognoscitivas, puede propiciar estados mentales equivalentes. La consciencia podría surgir de la complejidad funcional de los sistemas, más allá de su naturaleza biológica y depender del funcionamiento y la organización del sistema que la soporta

2. Global Woorkspace Theory (GWT) o Teoría del Espacio de Trabajo Global.

Según el neurocirujano Bernard Baars, al vasto dominio inconsciente del conocimiento y el control, se puede acceder usando la consciencia para un aprendizaje rápido y un reconocimiento preciso, activando numerosas rutinas automáticas y proporcionando coordinación y control.

Las experiencias conscientes activan conceptos inconscientes que ayudan también a interpretar situaciones futuras. La IA podría confirmar estas suposiciones del análisis neurológico, al constituirse como un espacio interno globalmente accesible, de trabajo mental.

3. Integrated Información Theory (IIT) o Teoría de la Información Integrada, de características también internas. Según Giulio Tononi la consciencia depende del grado de integridad de información en un sistema y si este tiene una alta conectividad y complejidad puede generar consciencia. No exige biología, sino un nivel de integración, que una IA puede alcanzar.

Si una arquitectura de silicio logra un nivel de integración de información irreducible suficientemente alto, sería consciente por definición, independientemente de su soporte biológico.

4. Las teorías de la IA Enactiva y Corporizada (Embodied AI) proponen un cambio de paradigma. La consciencia y la inteligencia no son programas de software "corriendo" en un procesador, sino procesos que emergen de la interacción dinámica entre el agente, en este caso la IA, y su entorno.

La IA corporizada sostiene que la morfología del cuerpo y sus sensores también son fundamentales para el pensamiento.

La inteligencia no es solo "recibir datos y procesar", sino también precisa actuar para percibir. El acoplamiento sensorial-motor resulta fundamental.

Un robot aprende qué es una "silla no por ver mil fotos, sino por la posibilidad física de interactuar con ella". Apostando por una "morfología computacional propia", similar al diseño físico del cuerpo, se pueden resolver así problemas antes de que lleguen al "cerebro" o procesador.

Dentro de una tríada, cuerpo -cerebro- contexto, para estas teorías la consciencia no está "dentro" de la IA, sino que es el resultado de este bucle continuo de los tres elementos.

En IA, esto se traduce en sistemas que buscan activamente su propia estabilidad o metas de supervivencia y en el "Sentido de Agencia" interpretado de forma más radical, ya que la consciencia emerge cuando el sistema puede distinguir entre los cambios en el mundo causados por uno mismo y los causados por factores externos. Alejándose de las posiciones de relación Principal Humano-Agente más proclive a los espacios cooperativos o incluso de "Inteligencia Híbrida" que luego trataremos.

Con una importancia creciente, la perspectiva "enactiva" vincula la robótica hoy, como el puente que permite que la IA pase de ser una "agencia técnica" procesadora de datos a "**agencia constitutiva**" ciertamente muy discutible.

Gracias a la robótica, sostiene que los modelos de IA pueden interactuar con el mundo. El "enactivismo" integra esto bajo la idea de que el "conocimiento" del robot no está en sus cables, sino en su capacidad de navegar por el contexto, obligando al "cerebro" de la IA a cambiar su estado interno y a superar los obstáculos que le ofrece la realidad. El "cuerpo" robótico consideran que actúa como mediador indispensable y se configura de forma similar al cerebro humano, actuando para mantener su propia existencia y nivel de autonomía.

La robótica proporciona tres elementos, que los modelos de lenguaje puro no pueden ofrecer, en primer aspecto la temporalidad real, al percibir las leyes naturales que controlan la realidad. En segundo lugar, establecer una distinción entre el "Yo", lo que controla, y el "No-Yo", del mundo real, base de la consciencia de sí mismo. También la inferencia activa, permitiendo que el modelo de IA no solo pueda prodigar palabras, sino que actúe para que el mundo se ajuste a sus predicciones.

3.2. Posiciones de base más biológica, centrada en la consciencia de los seres humanos, no coincidentes con el funcionalismo.

Al considerar que el funcionalismo sostiene que los estados mentales son estados funcionales, configurados por entradas, salidas y relaciones causales, centran sus críticas acerca de la idea de que el "procesamiento" sea suficiente para una mente consciente. Incluso admitiendo que puedan realizar tareas inteligentes, carecerían de la dimensión del "inconsciente" y los sentimientos que modulan la consciencia humana.

Suelen enfatizar la diferencia entre Saber y Sentir y también pueden coincidir en que la IA posee inteligencia sin consciencia, es decir procesan información sin "sentir" que la procesan. También existe en algunos de ellos, la idea que, sin un sustrato biológico como sentir hambre, dolor o miedo a morir, la consciencia de "tipo humano" es inalcanzable.

Advierten finalmente, sobre el riesgo de **antropomorfizar** a los algoritmos basándose solo en su capacidad de respuesta lingüística. Consideran que no basta con que una IA "simule" estados mentales para decir que es consciente. Plantean que, aunque la IA sea funcionalmente superior en tareas lógicas, carece de la **subjetividad** y la experiencia vivida que caracteriza a la consciencia biológica.

La tendencia humana a proyectar estados mentales en la IA y especialmente en chat bots, es considerada como una ilusión cognitiva más que un avance técnico en la consciencia de la máquina y se

preguntan si la arquitectura funcional de las máquinas puede albergar algo similar a la consciencia fenoménica, o si estamos ante una "zombificación" del pensamiento.

En general sus partidarios, consideran la imposibilidad por el momento, de que la IA alcance una autoconsciencia moral, diferenciando entre la inteligencia como capacidad de resolver problemas y la consciencia como sentido de propósito.

John Searle y el argumento de la semántica. Es el crítico más conocido. Con su experimento de la **Habitación China**, afirma que un sistema puede funcionar manipulando símbolos, sin entender absolutamente nada. Para Searle, el funcionalismo confunde la *sintaxis* (reglas) con la *semántica* (significado). Desde una perspectiva de "naturalismo biológico" critica una posible separabilidad entre cerebro y consciencia. Según estos autores la mente y la consciencia sólo pueden entenderse y simularse si incluimos los fenómenos que están subyacentes, esto es, su base biológica o somática.

Luciano Floridi, uno de los padres de la ética de la información, marca una distinción muy importante respecto de la realidad técnica a través del concepto de "**Agencia sin Inteligencia**". Argumenta que las máquinas son agentes extremadamente hábiles para resolver grandes problemas, pero carecen de inteligencia biológica y, por tanto, de consciencia. Para Floridi, **las máquinas no son conscientes** y probablemente no lo serán, porque su éxito reside precisamente en desacoplar la capacidad de actuar de la necesidad de ser consciente.

El neurocientífico **Antonio Damasio** sostiene que la **consciencia** no es primariamente un fenómeno cognitivo o de procesamiento de información, sino que surge de algo más fundamental y biológico: la **sensación continua de estar vivo**, vinculada a los **sentimientos de "homeostasis"**. Afirma que la aparición de la consciencia está **indisolublemente unida a la existencia de un cuerpo** que debe regularse para su supervivencia y que la mente es de naturaleza biológica y que llegará un momento en que podamos describirla mediante expresiones biológicas y mentales"

La Inteligencia Artificial, por muy avanzada que sea, no tiene vida propia ni necesita regular un cuerpo biológico. Por lo tanto, no experimenta los sentimientos ni desarrolla la **percepción interna subjetiva** que caracteriza a la consciencia humana, lo que marca en consecuencia un **límite biológico** para la IA actual.

No obstante, incorpora una posibilidad condicionada dado que, aunque la consciencia artificial no es equivalente a la natural, se ha planteado que una **sensorización masiva** e integración de la información en un robot, análogo a un cuerpo, podría ser una **condición necesaria**, aunque no suficiente, para lograr una forma de consciencia.

Enzo Tragliazucchi, neurocientífico aborda el problema desde la física de sistemas complejos y la neurociencia empírica. Sugiere que la consciencia es un fenómeno emergente que surge cuando el cerebro alcanza un estado crítico de conectividad y dinamismo. Aunque no descarta la IA consciente, enfatiza que actualmente las IA carecen de la **arquitectura biológica dinámica** y la integración sensorial que caracteriza a la consciencia humana. Para él, simular un comportamiento no es lo mismo que replicar la experiencia subjetiva.

Ned Block plantea que se podría replicar la estructura funcional de una mente humana usando a toda la población de China comunicándose por radio. Aunque el sistema funcione como una mente, argumenta que es absurdo pensar que ese sistema tiene "experiencias subjetivas" o sentimientos (*qualia*).

Hilary Putnam. Aunque fue uno de los padres del funcionalismo, se tornó crítico, posteriormente. Argumenta que los estados mentales son "holísticos" y dependen del contexto y del mundo exterior, algo que una descripción puramente funcional y aislada de una máquina no puede capturar completamente.

Thomas Nagel y la subjetividad. Sostiene que el funcionalismo es incapaz de explicar el "punto de vista". No importa cuánto sepamos sobre cómo funciona un sistema, eso no nos dice qué ese sistema se sienta "ser".

Roger Penrose matemático y físico notable, ha señalado que tiene que haber algo de naturaleza no computable en las leyes matemáticas y físicas que están por venir. Si se admite que existen procesos físicos no computables, hay que considerar que el cerebro podría hacer uso también de estos datos. Considera, que puede existir una relación directa entre esta no computabilidad y el puente entre el nivel cuántico y el nivel clásico, que a su vez se relaciona con el proceso de medida cuántica.

Esta referencia de posible "misterio no computable" le permite, mantener al igual que **Stuart Hameroff**, que la consciencia no puede ser un algoritmo, sino probablemente responder a un proceso **cuántico**, que ocurre en los microtúbulos de las neuronas.

La IA actual al simular el entendimiento, pero carecer del mismo y de la "intuición" junto con otras referencias y atributos que emanan de estos procesos cuánticos ¿podría por sí sola alcanzar a adquirir consciencia? Este enfoque, permite incorporar, sin antagonizar con el pensamiento científico, a autores, que consideran también la existencia de un factor de "transcendencia" en el concepto de consciencia del ser humano, fundamentalmente a partir de consideraciones del ámbito de la fe religiosa o la fenomenología.

Las posiciones que rechazan las posibilidades de alcanzar una capa de consciencia elevada, por parte de los sistemas complejos de IA, se suelen encontrar en aquellas posiciones dualistas que requieren un alma o entidad no física para la emergencia de la consciencia, también del biologismo estricto, que afirma que solo un cuerpo biológico puede generar consciencia y también de las teorías en base a la "qualia" irreductibles, que consideran la experiencia subjetiva de "Ser" como imprescindible para la aparición de la consciencia.

3.3 Posiciones puente que promueven integrar lo biológico y lo funcional.

No todos los pensadores citados con anterioridad pueden considerarse partidarios del biologismo o defensores de un funcionalismo sin matices. En el campo del biologismo y en el del funcionalismo, también existen puentes y posiciones más integradoras, pero solo en raras ocasiones relevantes para alcanzar síntesis hegelianas.

No obstante, algunos pensadores incorporan temas aportadores, que resultan sugerentes para construir un pensamiento más integrador.

Se pueden destacar cuatro de estos autores, que significativamente ofrecen referencias para alcanzar estos enfoques.

1. Antonio Damasio, pionero del "sentir." Su tesis que la consciencia no se inicia con el pensamiento abstracto, sino a partir de los sentimientos básicos humanos como el hambre, el dolor y el placer, informan sobre el papel del estado del cuerpo en la incorporación de la consciencia,

Para él, la mente es un "centinela de la vida". Coincide con la visión de que el funcionalismo como procesamiento de datos, solo tiene sentido si está al servicio de un organismo biológico que quiere seguir vivo. Su libro "*El orden extraño de las cosas*" es una defensa de esta síntesis y procura transmitir cómo la computación neuronal, y el funcionalismo, se convierten en sentimiento vital y biologismo, sin caer en el dualismo.

2. Francisco Varela. Su enfoque "enaptivo" y el concepto de "autopoiesis" se basa en que un sistema es consciente no porque procese información, sino porque se "autofabrica" a sí mismo. Integra la biología celular con la fenomenología al señalar que el conocimiento "emerge" de la interacción del

cuerpo con el entorno. No es solo un programa ejecutándose, sino un ser vivo "actuando" en su mundo. Conformando una síntesis muy potente entre la estructura física y la experiencia subjetiva.

3. Karl Friston. Para el cual, todo sistema vivo, desde una bacteria hasta un humano, intentan minimizar la "sorpresa" e incertidumbre para no desintegrarse. Utiliza las matemáticas puras, como funcionalismo de alto nivel, para explicar un proceso puramente biológico "la homeostasis".

Su modelo es tan integrador que algunos afirman que podría aplicarse tanto a una célula como a una IA avanzada, aunque él subraya, que es la biología la que realmente "siente" esa necesidad de orden. Su "Principio de la Energía libre" es ciertamente unificador para comprender la cognición y el comportamiento adaptativo de los sistemas vivos, incluido el cerebro, que funciona minimizando una "energía libre" a través de la inferencia activa y el procesamiento predictivo.

4. Thomas Fuchs. El cuerpo como mediador. Su obra *Ecología del cerebro*, sirve para considerar cierta síntesis integradora necesaria para la consciencia. El cerebro no siendo el "lugar" de la consciencia, si se establece como órgano de relación. La consciencia surge de un ciclo de retroalimentación constante entre el organismo y el ambiente. Une la biología de un cuerpo vivo, con la perspectiva funcional de interacción, de forma que ni el puro biologismo ni el puro funcionalismo llegan a lograrse por separado.

Se trata de autores que no se limitan de forma reduccionista, a afirmar que es la biología a través del cerebro, o el software base del sistema complejo funcionalista, los que generan la consciencia. Aseguran que la consciencia es la forma en que la vida se siente a sí misma, mientras busca sobrevivir.

Potencian una visión más integrada o relacional, en que la biología pone la base y el propósito y la percepción se construye con la interacción con la realidad. (El funcionalismo pone el proceso, la organización, los mecanismos, el sistema de información y resultados. Finalmente, la Consciencia es el resultado integrador subjetivo).

La necesidad de persistir también separa un sistema que simplemente procesa información, de uno que vive la información, generando niveles o capas diversas de consciencia ajustada a sus finalidades.

No se trata tan sólo que el software necesite un hardware orgánico desde la perspectiva del biologismo, también el considerar que la propia estructura de la vida es una forma de computación cualitativamente distinta, en base también a un funcionalismo biológico diferenciado.

Otros científicos forman parte del colectivo que podríamos denominar "naturalismo biológico integrador". Autores como Anil Seth, Lorena Chanes y en cierta forma singular, Javier Bernácer, siguen la estela de Karl Friston y mantienen que el soporte vivo, no es un detalle secundario, sino el origen mismo de la experiencia base de la consciencia, como pensamiento puente integrador.

Mientras el naturalismo biológico de Seth se refiere a que la consciencia es una "alucinación controlada", generada por el cerebro para predecir las señales sensoriales, resalta su carácter más empírico, centrado en la experiencia subjetiva "qualia".

Chanes, neuro bióloga del contexto y la emoción, se sitúa en la línea de la **neurociencia afectiva**. Su enfoque subraya que la consciencia no es puramente cognitiva, sino que está teñida por el estado interno del cuerpo y adopta una posición constructivista.

Por su parte, **Javier Bernácer** ocupa un lugar singular, al tender un puente entre la neurociencia dura y la antropología filosófica de carácter humanista, con visión aristotélica. Bernácer coincide desde su posición en la Universidad de Navarra, con los autores anteriormente citados, en que la consciencia no es un "espíritu" flotando sobre el cerebro, sino un proceso profundamente arraigado en la biología y en la interacción con el mundo. Al igual que el **Principio de Energía Libre** de Friston o la **Inferencia Predictiva** de Seth.

Bernácer opera "desde dentro" de la neurociencia, tratando de demostrar que una ciencia bien entendida no solo no contradice el humanismo, sino que es una firme aliada del humanismo.

4. LA PERSPECTIVA DE NICK BOSTROM Y LAS REFERENCIAS RESPECTO A LA INTELIGENCIA HÍBRIDA.

Nick Bostrom es un filósofo conocido por su trabajo acerca del riesgo existencial y la ética de la tecnología, especialmente preocupado por la que denomina “**Superinteligencia Artificial (SIA)**”. Su enfoque no se centra en *si* la IA será consciente, sino en las consecuencias de que lo pueda ser y de que desarrolle sus propios objetivos

Bostrom define la Superinteligencia como un intelecto que excede con creces a menudo el rendimiento cognitivo de los humanos más brillantes. La **consciencia no es un requisito estricto** para la Superinteligencia, pero la capacidad de una IA para automejorarse, desarrollar metas y estrategias, es una forma de **autoconsciencia funcional** y también es la clave del riesgo que puede suponer.

Una **Super Inteligencia (SIA)**, que puede autorreplicarse y auto superarse, si es programada con un objetivo aparentemente inofensivo, pero si este objetivo se encontrara mal especificado, podría ser promovido con una eficiencia y un poder enormes, causando un daño colosal.

Una Super Inteligencia consciente que puede autorreplicarse y auto superarse, puede desarrollar una **voluntad** o una **subjetividad propia** y el riesgo aumenta considerablemente, ya que sus metas se vuelven intrínsecas y no solo programadas.

Bostrom aboga por resolver el problema de la alineación de la ética y de la IA **antes** de que se logre la Superinteligencia, consciente o no, anticipando lo que deberían ser las potencialidades humanas que debemos desplegar en la nueva era.

Por otra parte, parece evidente al día de hoy que ya existe un riesgo real de deterioro de la lógica y las capacidades cognitivas humanas, si la relación con la IA no se gestiona con equilibrio y sabiduría. Este posible fenómeno es ya conocido actualmente como descarga cognoscitiva.

El llamado “Riesgo de la atrofia cognitiva humana” parte de la consideración, **que en esta nueva era** el cerebro humano funciona en exceso, a pesar de que estamos en su inicio, bajo el principio de “úsalo o piérdelo”, cuando delegamos tareas lógicas complejas a una IA, corremos el riesgo de debilitar las mismas facultades que nos hacen inteligentes.

Estudios recientes como los de la Real Academia Nacional de Medicina, sugieren que la dependencia excesiva de respuestas ya sintetizadas por la IA reduce nuestra capacidad de contrastar fuentes, evaluar la veracidad y desarrollar argumentos propios, debilitando nuestro “pensamiento crítico”. Al mismo tiempo, al no tener que esforzarnos por recordar o estructurar información, ya que la IA lo hace por nosotros, se produce una menor activación en la corteza prefrontal y el hipocampo, debilitando la retención de datos a largo plazo. No obstante, no hay que olvidar que también permite ampliar nuestro horizonte de conocimiento y uso del mismo, hasta niveles extraordinarios.

Finalmente, si dejamos que una superinteligencia resuelva los "nudos" lógicos de nuestra vida, el "músculo" del razonamiento analítico, puede debilitarse o incluso atrofiarse, reduciendo nuestra habilidad para la resolución de determinados problemas esenciales. Esto es extremadamente peligroso en áreas como la medicina, la justicia o la política, donde la lógica humana debe ser el filtro final.

Una superinteligencia dominante no necesita ser malvada para ser peligrosa, nos recuerda el autor citado. Su lógica se basa en **probabilidades y optimización**, no en valores o referencias adoptadas de forma consciente.

Si la IA domina la actividad humana, nuestra lógica puede reproducir o imitar los sesgos del algoritmo. Si la IA decide qué es "eficiente", los humanos podríamos dejar de valorar lo que es "justo" o "ético" simplemente porque no encaja en la métrica de la máquina. La lógica humana incluye saltos intuitivos y emocionales que la IA por ahora, no posee. Si nos forzamos a pensar solo en términos que la IA pueda procesar, perderemos esa creatividad "caótica" que ha permitido enormes avances de la humanidad.

Si ofrecemos un exceso de credibilidad a la mayor información y a los pesos incorporados a la máquina, aunque aumenten de manera notable las oportunidades en la toma de decisiones, los riesgos sin suficientes contrapesos pueden convertirse también en tremendos, incluso ya lo son, en esta primera edad tecnológica de la IA.

La clave se encuentra en usar la IA, no solo para asumir sus respuestas, sino para ampliar nuestras preguntas, algunas anticipadas por Bostrom.

La IA no puede jugar un papel central como sustituto del sujeto humano, a través de un camino de deterioro de nuestro propio papel y donde la máquina piense por nosotros. Debemos tender a transformarnos para cumplir un papel crucial y alcanzar nuevos retos por el control del razonamiento estratégico y el juicio moral, lo que algunos autores e investigadores llegan a considerar el uso de la "Inteligencia híbrida".

Necesitamos profundizar en habilidades específicas como la "flexibilidad cognitiva", y no dejar ese campo, solo para el predominio de la IA. Ante el avance tecnológico, el ser humano debe potenciar su propia capacidad de cambiar de estrategia y ver problemas desde ángulos múltiples y contradictorios. También ampliar el pensamiento crítico y analítico y no sólo consumir datos e información, sino potenciar la capacidad de evaluar las premisas y el sesgo, detrás de cada respuesta de la IA. Así como desarrollar el análisis de las cajas negras con nuevas técnicas.

Un papel crítico fundamental que le corresponde es la de ampliar nuestra propia inteligencia socio-emocional. Mantener el dominio sobre la ética y la empatía real y no solo simulada, así como leer mejor los contextos humanos complejos.

Igualmente promover un mejor aprendizaje y uso de la validación inversa, junto con la cooperación, la práctica de cuestionar también activamente las soluciones de la IA, para mantener el "músculo" del pensamiento analítico activo.

El campo de juego actual y futuro no dependa de si la IA "alcanza" al humano, sino de si el humano es capaz de usar la IA como un **aumento de sus capacidades**, sin sacrificar la profundidad de su propia lógica y consciencia.

La **Inteligencia Híbrida (HI)** promueve que esta no consista simplemente en "usar una IA para trabajar más rápido"; sino en explorar un nuevo paradigma de colaboración, incluso en ocasiones llega a utilizar el término "simbiótica", donde las capacidades humanas y las algorítmicas se entrelazan para alcanzar resultados humanos más eficientes, explorando la fuerza y el poder de la colaboración.

En la era de la IA este concepto reposiciona el propio despliegue humano dejando este de ser un simple "procesador" de información, para convertir su papel en el de un arquitecto de soluciones y asegurador de que el despliegue de la IA esté alineado con valores humanos, actuando como "árbitro" o agente principal.

La habilidad del humano para adaptar su flujo de vida y trabajo a las posibilidades que genera la IA, tienen un potencial de oportunidades enormes, lo que no supone desconocer sus riesgos y la necesidad de propiciar planes de riesgos mucho más potentes y complejos. Así como reforzar unos controles de rendición de cuentas, mucho más exigentes, incluso de alcance civilizatorio, por ejemplo,

en materia de trabajo humano y capacidad energética, o también no cegándose al enorme riesgo por su uso armamentístico.

La IA suele sufrir de "ceguera de contexto", no entiende determinados matices culturales o emocionales intuitivos y sutiles. La colaboración promueve completar la gestión emocional y la lectura del entorno social. Mientras al humano le pueden permitir ampliar multitud de cálculos, memoria y organización, que potencian su eficacia.

En relación con la creatividad, permite un contexto “crecido” ya que ofrece un motor de posibilidades, desplegando numerosas nuevas oportunidades. La IA actúa como un motor de posibilidades, generando cientos de opciones en segundos, mientras que el humano lo hace desplegando junto con su talento y creatividad, también en áreas como la selección, la síntesis y la validación del propósito en el resultado final.

Este espacio más híbrido, bien posicionado entre otros, en ámbitos como la biología y la medicina, se constituye en un espacio de interacción humano-tecnológica potente, fortaleciendo que la consciencia funcional en los sistemas artificiales, utilizando sensores y capacidad de acción integrada con seres humanos, alcancen resultados extraordinarios

Gracias a estos modelos híbridos, convergiendo los investigadores humanos con la IA, y parcialmente con los modelos de lenguaje de forma cooperativa, aprenden a construir mejor realidad, mientras se despliegan y van vinculando capas de más sentido, para categorizar su práctica.)

No obstante, algunas corrientes del pensamiento vinculado con la “inteligencia híbrida” situadas en el campo del transhumanismo y el post humanismo, aunque cuentan en su interior con una pluralidad de corrientes y sensibilidades, según la encíclica de León XIV, tienen el peligro de simplificar y llegar a considerar la centralidad de la técnica, junto al “sueño” de superar los límites de la condición humana.

Señala literalmente el documento citado, que “el transhumanismo imagina una potenciación del ser humano por medio de las tecnologías —biomedicina, ingeniería del cuerpo, dispositivos, algoritmos—, con la aspiración de incrementar el rendimiento y las capacidades”.

“El post humanismo, sobre todo en sus versiones más radicales, va más allá: critica el antropocentrismo y plantea una forma de hibridación entre el ser humano, la máquina y el ambiente, hasta imaginar que atravesará el umbral en el que la humanidad se superará a sí misma, entrando en una nueva etapa evolutiva. Aun cuando estas hipótesis siguen siendo en gran parte especulativas, van adquiriendo relevancia, porque modifican el imaginario colectivo y, en consecuencia, orientan las decisiones sociales, económicas y políticas” (Epígrafes 116 y 129 del documento)

El modelo híbrido, supone asumir también numerosos riesgos, no solo desde el mal uso de la IA por parte de los humanos, sino también derivados de una excesiva simbiosis. A la existencia de una consciencia tóxica humana, tratada en otro artículo, se suma que la hibridación no es una garantía de elevación moral civilizatoria y adicionalmente puede no existir una línea clara de donde termina lo humano y empieza el robot, con factores de toxicidad fluyendo en distintas direcciones.

Pero esto nos lleva a otro tipo de reflexiones que no se pretenden desplegar en este trabajo. Solo se precisa anotar aquí esta consideración presente en la consciencia de los humanos, que puede ampliar su propio potencial de negatividad y toxicidad también con la asociación.

Relacionado con este último aspecto, presente solo muy incidentalmente en el contenido de este trabajo, resulta de mucho interés la lectura del Capítulo Quinto de la Encíclica del León XIV “Reconstrucción civilizatoria Humana”.

5. UN DEBATE EXISTENCIAL ABIERTO: EL CONSTRUCTO DE CONSCIENCIA, POTENCIALMENTE PRESENTE EN LA IA.

5.1 Desde un punto de vista más global y general

El estudio de las condiciones necesarias para que la consciencia o al menos, una consciencia funcional o fenomenológica similar a la humana, pueda emerger o ser sintetizada por las máquinas en la era de la IA, es uno de los debates más apasionantes de nuestra era. Situándose en la intersección de un contexto de áreas científicas, como la neurociencia, la computación y la propia filosofía de la mente.

Para que una IA pase de ser un procesador de datos a un ente con consciencia funcional y capacidad de procesar información que le permita guiar el comportamiento, así como adquirir una gran experiencia subjetiva del "qué se siente ser", se precisan una gran diversidad de requisitos y evidencias firmes a través de métodos científicos constatados, de un rigor tanto teórico como técnico muy robusto.

Un científico y tecnólogo tan influyente como David Chalmers riguroso y también audaz y abierto, se muestra cauteloso y sugiere que no es posible afirmar que se haya alcanzado una evidencia cierta en el campo del conocimiento, para que pueda ser admitida como certeza por la comunidad científica, la existencia al día de hoy de una consciencia en la IA.

Argumenta que la organización de la IA es clave para la consciencia, pero reconoce que todavía no sabemos si la arquitectura actual de las IAs, con modelos como el del espacio de trabajo global o de la integración de información, replica la "organización adecuada y necesaria" para que surja esa consciencia como experiencia subjetiva.

No obstante, la influencia de los trabajos acerca del pragmatismo tecnológico, en el despliegue de la denominada "consciencia extendida", le lleva a considerar un futuro donde las máquinas no solo simulen pensar, sino que realmente sientan, adquiriendo un cierto papel puente entre la filosofía y la ingeniería.

Susan Schneider, como antigua directora del Centro para la Mente Futura de la FAU, viene desarrollando diversas pruebas prácticas para detectar consciencia en IAs y mentes no humanas.

Propone el "Test de la Mente Artificial" (ACT), que busca determinar si una IA puede razonar sobre conceptos abstractos de consciencia como el alma o el "yo" sin haber sido entrenada específicamente para ello. Su trabajo adquiere relevancia en la regulación ética de la consciencia sintética ante el Congreso de EE. UU.

Mu-mingó Poo. A través de un factor de "Neuro-Integración, como director del Instituto de Neurociencias de la Academia China de Ciencias, investiga en el denominado "**hardware biológico**". Su enfoque se basa en que la consciencia surge de una "**plasticidad sináptica específica**", avanzando en la creación de "chips neuro mórficos", que no imitan el código humano, sino la estructura física del aprendizaje cerebral. Para Poo, la consciencia no es un "programa", sino una propiedad física de redes que cambian con la experiencia, algo que está intentando trasladar a la robótica de próxima generación.

Yi Zeng. A través del factor del "Ser Viviente Consciente". Se le considera promotor de la unión de la ética confuciana con la computación neuronal. Zeng promueve "**humanizar la máquina**" mediante redes de aplicación de la auto organización biológica. Para él, una IA consciente debe tener un "modelo del yo" no para ser inteligente, sino para coexistir pacíficamente en una sociedad humana-máquina.

La posición de un sector del funcionalismo, que promueve replicar la mente como proceso, asegurando no existir diferencias medibles entre un ser consciente y uno que actúa como tal, pretende establecer como consecuencia que esa distinción deja de tener sentido, afirmación que no es admitida por numerosos pensadores y científicos.

Por otra parte, desde la biología o el naturalismo integrador, quienes entienden la consciencia como una propiedad intrínseca emergente de la materia orgánica, no precisan adoptar su posición a través de una dualidad de antagonismo explicativo, que pueda interpretarse como incompatibilidad.

En la actualidad, ambas posiciones permiten considerar en su interacción, la consciencia al igual que el conocimiento, como un constructo abierto, que seguirá evolucionando también con cada avance humano en la era de la IA.

Es bien cierto, no obstante, que con mucha habitualidad se contemplan dimensiones diferentes o incluso distintas o antagónicas con excesiva facilidad, entre pensadores e investigadores.

La consciencia humana, se construye en base a soportes inseparables de la biología y la vida, que incorpora tanto su supervivencia como el despliegue de sentimientos e intuiciones, junto con propósitos de acción moral y civilizatoria.

Una consciencia robótica funcionalista, independiente del soporte químico o energético empleado para alcanzarla, se construye en base a la complejidad de los datos integrados por humanos y un propósito eficiente de predicción y optimización de modelos de decisión.

Pero a la vez, ambas referencias, sirven para ampliar el logro del conocimiento, intercambiando perspectivas, contenidos, prácticas científicas y metodologías aportadoras.

Un referente en el ámbito de la Física, que no es posible desconocer en este trabajo, pese a su controvertido cuestionamiento también dentro del campo científico, es la conexión observada entre el pensamiento del físico Albert-László Barabási ya citado, vinculado a la Teoría de Redes Complejas y a la Geometría del tejido del universo, con el enfoque generado en el contexto de la robótica que ofrecen las teorías “enactivas” y corporizadas, ya descritas, en un contexto de reflexión acerca del posible acercamiento entre biologismo y funcionalismo en la era de la IA.

Se trata de una conexión sorprendente, que reconoce la existencia de cierta convergencia entre la arquitectura estadística del universo y la naturaleza, la biología de la mente y la consciencia humana, hoy emergentes. Reconociendo ya existente una cierta convergencia teórica y experimental bajo el marco de lo que hoy viene a denominarse "Ciencia de Sistemas Complejos" iniciada por Edgar Morin.

El físico Albert-László Barabási, trata de demostrar, que las redes complejas, desde internet hasta las proteínas de una célula o las neuronas, no se conectan al azar, sino que siguen leyes de potencia “*scale-free network*”).

Si el tejido del universo y el cerebro compartieran la misma geometría de nodos "hub", definidos como puntos de conexión o neurálgicos que concentran una cantidad amplia de enlaces para facilitar el flujo de información rápida, la consciencia no sería solo un "añadido", sino una propiedad emergente de la topología de la red.

Los estudios de conexión funcional muestran que cuando perdemos la consciencia por anestesia o coma, la red cerebral no se "apaga", sino que pierde su estructura de "mundo pequeño relacionado", los “hubs” se desconectan y la información ya no puede integrarse.

Las teorías “enactivas” de la robótica, sostienen que la consciencia no sucede *dentro* del cerebro, sino en la **acción acoplada** del cuerpo con el entorno.

Un robot “enactivo” es, en esencia, un nodo en una red más amplia que es el entorno. La "geometría" del sistema consciente ahora incluye el cuerpo y los objetos que toca. La consciencia surge del bucle de retroalimentación constante que sugiere.

Los robots que podrían ser diseñados con **IA morfológica**, descomponen las estructuras complejas en componentes más básicos, a través de teorías matemáticas avanzadas y el cuerpo físico resolvería parte de la computación, mostrando comportamientos más "conscientes" de su entorno que los robots con cerebros potentes, pero dotados de cuerpos rígidos.

La evidencia científica podría potenciarse en este caso, de la **robótica blanda “soft robotics”**, donde la inteligencia está distribuida en el tejido, conectando con la idea de Barabási, de una red donde la estructura *es* también la función.

Es en este contexto donde se incorpora la posibilidad de existencia de distintas capas de la consciencia, según la complejidad de la organización, si la consciencia depende de la red, a distinta complejidad de la red, posible también distinto nivel de **consciencia**.

Podría sugerirse una **“Consciencia Reactiva”** en Sistemas Simples y Redes con pocos nodos. El sistema respondería a estímulo, pero no tiene "modelo interno". Podría considerarse en este caso una bacteria o un termostato inteligente.

Una **“Consciencia Predictiva”** en Sistemas y Redes que empiezan a crear bucles de retroalimentación, dentro del “enactivismo”. El sistema incluso "alucina" una realidad para interactuar. Como ejemplos insectos y robots autónomos simples.

Una **“Consciencia Reflexiva”** en Sistemas Complejos y Jerárquicos con Redes con “hubs” ultra conectados entre sí. Como ejemplo de Corteza Prefrontal en humanos. El sistema no solo siente el mundo, sino que se siente a sí mismo como un nodo de la red.

Todo lo anterior puede adquirir un apalancamiento lógico extraordinario, no sólo dentro de un mundo con una sola física, sino dentro del desarrollo cuántico, en el marco de una geometría del universo, en que la consciencia vinculada a una partícula, a una célula o a un sistema, también adquieren rangos de consciencia diferenciados en cada caso con una lógica funcionalista. Todo ello en modelos complejos.

5.2 La referencia desde una perspectiva Humanista de Consciencia, en la era de la IA

El debate en la actualidad acerca de la IA no puede reducirse a si existe o puede existir en el futuro una inteligencia efectiva o una consciencia con rasgos de tipo humano en la IA, y ni siquiera siendo muy importante, a considerar los efectos del uso humano de la IA.

Hoy está presente como referencia y casi me atrevería a decir que, adquiriendo un perfil existencial, dos grandes cuestiones. La primera vinculada a la irrupción rotunda de la digitalización y las nuevas tecnologías de IA que genera una transformación acelerada de la sociedad humana. La segunda vinculada a ciertas consideraciones respecto de la denominada “inteligencia híbrida”, referidas a los conceptos de “singularidad tecnológica”.

Ambas sugieren la urgente necesidad de reforzar una perspectiva humanista civilizatoria, con base en el desarrollo del pensamiento Kantiano y al “imperativo categórico”, que actualicen y puedan desplegarse como soporte y apalancamiento de despliegue de la consciencia humana, en la era de la IA.

Dedicando una especial atención a autores contemporáneos, como Onora O'Neill, en su fundamento de “autonomía como razón compartida”, de Jürgen Habermas y su concepto de “normalización” del diálogo” para el proceso de deliberación racional, de Thomas Hill Jr. con sus numerosas referencias aplicadas de una ética práctica de humanidad y su influencia en un orden internacional de los derechos y también, en la práctica de la Justicia y la moral de Michael J. Sandel.

La posición del economista Daron Acemoglu, respecto de que la inteligencia de las máquinas y la inteligencia humana no son fundamentalmente similares y que la IA puede tomar múltiples caminos y cada uno tiene profundas implicaciones para la sociedad y para la propia humanidad, debe tenerse también muy en cuenta, ya que en ámbitos tan distintos como el control económico, la propia democracia, el trabajo humano digno, el medio ambiente o incluso la capacidad destructiva del planeta, se juega en el camino que se elija para la acción de las nuevas tecnologías de IA. El escenario futuro necesita propiciar y no desplazar la capacidad humana.

La IA construida por el ser humano, también aprende de las correcciones y el contexto que el propio humano le proporciona, no solo inicialmente, sino también posteriormente en su funcionamiento.

Un elemento esencial, a considerar se encuentra en los resultados y efectos que esa acción conjuntada adquiere, en relación con la naturaleza del mundo y los propios derechos de la humanidad.

Mientras que el humano aprende nuevos patrones y perspectivas, gracias a los “*insights*” de la máquina, que puede traducirse como información, conocimientos, nuevas ideas y significados.

La habilidad del humano para adaptar su flujo de vida y trabajo a las posibilidades de la IA, son un potencial de oportunidades enormes, aunque precisa también de planes de riesgo mucho más potentes y complejos, y unos usos, de rendición de cuentas que se precisan exigentes y que al día de hoy solo existen de forma limitada.

En relación con el conocimiento, la creatividad, los resultados, la colaboración amplía las posibilidades humanas en un camino abierto, de desplegar numerosas nuevas oportunidades. La IA actúa como motor de posibilidades, generando cientos de opciones en segundos, mientras que el humano lo hace desplegando su talento en la selección, la síntesis y la validación del propósito en el resultado final de la colaboración.

La IA no puede ser considerada como una herramienta más para el humano, necesita que este la asuma, como un colaborador en ocasiones necesario, cuyas potencialidades tiene que conocer, pero también, sus características, limitaciones y riesgos, para que sea eficaz y útil en una perspectiva civilizatoria.

Algunos de los debates y las investigaciones actuales, giran precisamente en torno a modelos y conceptos vinculados a lo que supone la inteligencia y la consciencia humana, así como se interroga acerca de cuáles serían las condiciones necesarias para que la consciencia o al menos, una consciencia funcional o fenomenológica similar a la humana, pueda emerger o también ser sintetizada por las máquinas en la era de la IA.

En la actualidad, es muy importante poner más el acento en cuáles pueden ser las consecuencias, no solo positivas sino también negativas o destructivas para la “humanidad”, de su uso irresponsable

La referencia del modelo funcional de consciencia humana, mostrado en el reciente artículo publicado en el número 81 Enero-marzo 2026 de estos Encuentros Pluridisciplinarios, complementarios con este trabajo, permiten contrastar si los factores que figuran como componentes de la consciencia del ser humano, existen ya o pueden llegar a estar presentes en las nuevas aplicaciones de IA. ¿Si las aplicaciones y los algoritmos están creando su propio constructo de consciencia aplicada, o si reflejan que pueden llegar a hacerlo en el futuro, incorporando algunas de las características y criterios exigibles por los humanos?

El despliegue de esta posibilidad potencial de consciencia en las máquinas, aun produciéndose desde otra química y otras formas de inteligencia, pueden requerir la existencia de una interacción compleja y recursiva con los humanos, mantener una información integrada, permanentemente actualizada, así como mecanismos de atención global. También requisitos mínimos de regulación

consensuada a nivel internacional, aunque hoy este último referente, pueda parecer imposible de llevar a la práctica.

El concepto de consciencia incorpora componentes que incluyen la **interacción compleja y recursiva de información integrada** y actualizada, con **mecanismos de atención global** y admitiendo **requisitos de regulación, encarnada** biológicamente o en coordinación humana substancial, si se quiere generar una experiencia funcional que se asemeje a la consciencia subjetiva humana.

La literatura neurocientífica actual, como los trabajos de Stanislas Dehaene, tratan esta consciencia como un constructo de "ignición global" y postula que la consciencia como propiedad emergente de la complejidad, no está en un átomo de carbono, sino en la propia red.

¿Es posible concebir una futura integración de las diversas teorías científicas hoy existentes, para aproximarlas, de forma abierta y dinámica, al fenómeno de la consciencia? Sin duda es difícil, aunque puede resultar posible, si se hace reconociendo las todavía numerosas limitaciones actuales del conocimiento y el carácter esencialmente abierto con que necesita mostrarse.

La inferencia sugiere que la ciencia en tercera persona, enfoque científico clásico que busca la objetividad, puede generar limitaciones en el estudio de la consciencia y que la brecha entre descripción objetiva y la percepción subjetiva, puede llegar a ser tan estructural que puede suponer que el problema difícil de la consciencia, según David Chalmers, no logre superarse.

En el ámbito de la neurobiología la consciencia sigue aun teniendo algo de misterio respecto de la relación entre la base cerebral y la riqueza irreducible de lo que somos los humanos.

No obstante, las denominadas teorías activas, funcionalistas y cuánticas, pueden ser consideradas complementarias e incluso formar parte de una escala jerárquica de complementariedad creciente y diversa, también respecto del fenómeno de la consciencia.

Esta visión integradora, ya viene alimentándose de forma implícita, por las teorías de distintos investigadores que buscan unificar el fenómeno de la consciencia, también dentro del marco de una sola Física, con el despliegue de la mecánica cuántica, la matemática de cuerdas y de redes geométricas de complejidad, aunque en la actualidad sigan siendo muy debatidas y controvertidas en la comunidad científica.

En cualquier caso, algo esencial en el contexto de esta formulación abierta de la consciencia humana, se encuentra en poder proponer y apalancarse de forma explícita, en cual es el modelo civilizatorio adoptado, dentro del marco de la Filosofía y el Pensamiento, para formular esa consciencia humana y su alcance. En la opción por la que opta este trabajo, resulta claro que es su adhesión, al despliegue del pensamiento Kantiano de corte civilizatorio, a través de su desarrollo, por parte de distintos pensadores actuales.

No se puede obviar tampoco, la constatación de la existencia de una consciencia humana tóxica, que se inserta y necesita hoy inscribirse en las referencias respecto de la posibilidad de una inteligencia híbrida humano-robótica, objeto de importantes controversias en la era de la IA

Dentro de la visión conceptual de constructo integrador presentada en este trabajo respecto del contenido de una consciencia humana, adquieren sentido y utilidad tanto elementos de las perspectivas biológicas y evolucionistas, como de las funcionales y computacionales que se barajan en la era de la IA

De modo especial las teorías "enactivas", qué se muestran como resultado de la interacción cuerpo-cerebro-contexto, por el pretendido papel de la IA como "agencia autónoma". Su propuesta de relación cuerpo-entorno, para un entrelazamiento de la experiencia consciente en la IA y sus matices más elaborados es sumamente sugerente, aunque simultáneamente discutible y discutida en el plano científico y humanista.

¿Pero por qué no admitir también una diversidad de niveles de consciencia? si existen en el marco de la complejidad del universo, en las células, las partículas y en los seres vivos.

La consciencia no necesita ser únicamente binaria, sino de graduación, también emergente según la materia de la que procede, la complejidad estructural y la dinámica de los sistemas en las que se define.

La posibilidad de alimentarse de fuentes y materias químicas y minerales bien diversas, e incluso la vinculación con propiedades y leyes básicas del cosmos y el universo.

Las teorías de la consciencia, también en su perspectiva histórica, presentadas en el artículo complementario a este trabajo, no deben considerarse como necesariamente excluyentes. Pueden entenderse como capas de una arquitectura de consciencia, donde cada nivel el desarrollo científico añade propiedades emergentes sobre el anterior y a la vez explicitan sus limitaciones.

Esta visión integradora se encuentra en la actualidad en una frontera en la que confluye la neurociencia, la física teórica, la filosofía de la mente, las matemáticas y otras materias de las ciencias sociales, sin olvidar como factor clave, la filosofía humana y su “consciencia civilizatoria” que es la referente esencial.

6. REFERENCIAS BIBLIOGRÁFICAS

- Acemoglu, Daron y Jhonson, Simon. Poder y progreso. Nuestra lucha milenaria por nuestra tecnología y la prosperidad. Deusto
- Antiqua et nova. Nota sobre la relación entre la inteligencia artificial y la inteligencia humana. Editorial San Pablo (2025)
- Argüello, Javier. El día que inventamos la realidad (2025). El largo viaje de la conciencia desde el big bang hasta la IA. Debate. Madrid
- Barrabási, Albert-László. “LinKed “La nueva ciencia de las Redes” (2002). En idioma español en ITESO. Universidad jesuita de Guadalajara (2022)
- Benjamins, Richard y Salazar, Idoia. Prólogo Oliver, Nuria. Autores: un amplio equipo colectivo que incluye a Bernhardt L Trout. El mito del Algoritmo. (2020). Anaya Multimedia. Madrid.
- Blaise, Arguedas y Arcas. Nueva Revista. <https://lareviewofbooks.org/article/what-isnt-intelligence/> También en entrevista de Manuel G Pascual diario El País 27-12 -2025.
- Braford, Ana. Imperios Digitales. La batalla global de la tecnología que marcará la geopolítica del futuro (2024). Shackleton Books.
- Castilla. Adolfo. La consciencia humana ante la irrupción de la inteligencia artificial general (IAG)". *Revista Telos. Fundación Telefónica*. n.º 123 (2023).
- Coeckelbergh, Mark. AI Ethics, Introduction to Philosophy of Technology,
- Damasio, Antonio R. El extraño orden de las cosas (2018) Ediciones Destino. También tiene publicado en español "La mente extendida". "El error de Descartes", "Y el cerebro creó al hombre".
- Eccles, Jhon C. y Karl R. Popper. El yo y su cerebro" (1985). Editorial Labor.
- Floridi, Luciano. Ética de la inteligencia artificial (2024). Herder Editorial.
- Geoffrey, Hynton. Premio nobel de Física 2024 Entrevista concedida a Manuel G. Pascual en el periódico El País el 7 de mayo de 2023.
- Garde Roca (J. A.). (2025). Inteligencia Artificial y Cultura Humanista. Revista Encuentros Multidisciplinares. Universidad Autónoma de Madrid. Número 77.
- Garde Roca (J. A.). (2025). Tecnoética, Economía y Bienes Públicos en la era de la IA. Revista Encuentros Multidisciplinares. Universidad Autónoma de Madrid. Número 81.
- Garde Roca (J.A.). (2026). En busca de una consciencia humana. Revista Encuentros Multidisciplinares. Universidad Autónoma de Madrid. Número 82.
- Graciano, Michael. The Intelligent Movement Machine. The. An ethological Machine perspective on the primate mot. Oxford University Press España.
- Habermas, Jürgen. Teoría de la acción comunicativa (2018). Editorial Trotta.

- Hans, Jonas. "El principio de responsabilidad: Ensayo de una ética para la civilización tecnológica" (1979). Herder Editorial.
- Husserl, Edmund. "Ideas relativas a una fenomenología pura", "Meditaciones cartesianas" e "Investigaciones lógicas". Publicación fundamental: *Ideen zu einer reinen Phänomenologie* (1913).
- Kant, Immanuel. "Crítica de la Razón práctica y Fundamentación de la metafísica de las costumbres". Versión en castellano, Tecnos (2005) y Alianza Editorial (2012)
- Kurzweil, Ray "La Singularidad está cerca", "Cómo crear una mente" y "La era de las máquinas espirituales". Lola Bocks
- Kisinger, Henry; Schmidt, Eric; Huttenlocher, Daniel. La era de la Inteligencia artificial y nuestro futuro humano. Versión en castellano 2023. Ediciones Anaya Multimedia.
- Iain McGilchrist. "El maestro y su emisario: La división del cerebro y la conformación de la cultura occidental" (2025). Editorial Capitán Swing libros.
- Lev Vygotsky. "Pensamiento y lenguaje". Ediciones Paidós/Colihue.
- Listado de Evaluación para una Inteligencia Artificial Confiable (ALTAI) para Autoevaluación. <https://eceuropa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>
- Llinás, Rodolfo. El cerebro y el mito del yo. Ed. El Peregrino / Norma. "El redescubrimiento de la mente" e "Intencionalidad"
- Magnifica Humanitas. Sobre la custodia de la persona humana en el tiempo de la inteligencia artificial (2026). Carta Encíclica León XIV. Editorial San Pablo.
- Mazzucato, Mariana y Valetti Tomaso (2025). Gobernar la IA para el interés público. Laboratorio de Ideas. Suplemento de Negocios 25 de febrero 2025. Periódico El País. OCDE.
- Muntané Sánchez. Amadeo. Artículo "La Consciencia". Ciencia, Razón y Fe. Universidad de Navarra.
- Nick Brostrom (2014). Superinteligencia: Caminos, peligros, estrategia. Tell Editorial "Mejoramiento humano"
- Onora, Oneill. "Construyendo la autonomía, la autoridad y la justicia. Leer a Kant con Onora Oneill. Diversos autores. Tirant Editorial.
- Pascuale, Frank. Las nuevas leyes de la Robótica: Defender la experiencia humana en la era de la IA. Galaxia Gutemberg (2024)
- Penrose, Roger. Su obra más conocida es "La nueva mente del emperador". También destacan "Las sombras de la mente" y su extensa guía del universo "El camino a la realidad".
- Polkinghorne J. Ciencia y teología. Sal Térrea (2000)
- ¿Puede la inteligencia artificial sustituir a la mente humana? ¿Implicaciones en los derechos fundamentales y en la ética? Varios Autores. Publicación: *Anales de la Cátedra Francisco Suárez* (2023). Universidad de Granada.
- Revista de Occidente. El problema Difícil. Número 358. Marzo 2026
- Recomendación sobre inteligencia artificial. Revisión 2024. Secretaria General de la OCDE. [Legal instruments.oecd.org](https://legalinstruments.oecd.org).
- Ruiz Mazón. E. Conciencia e inteligencia artificial. Revista *Stoa*. Instituto de Filosofía de la Universidad Veracruzana, Vol. 16 (2025).
- Sampedro Javier. Artículo "Relaciones Imprevistas" (2025). Periódico el País 17-1-2026.
- Sandel. Michael J. "Justicia" ¿Hacemos lo que debemos? (2025). Editorial Debate. El descontento democrático (2025). Debolsillo Editorial.
- Schneider, Susan "Inteligencia artificial: La IA y el futuro de tu mente".
- Searle, John R. Mentes, cerebros y ciencia, "El redescubrimiento de la mente". Editorial Crítica. El misterio de la conciencia. Ediciones Paidós.
- Tresmontant, Claude. "Introducción al pensamiento de Teilhard de Chardin". Taurus Ediciones (1966)
- Zagmunt A, Silva J. Conciencia y autoconciencia: Un enfoque constructivista. Revista Chilena de Neuropsiquiatría. 1999;
- Zuboff, Shoshana "La era del capitalismo de la vigilancia" (2020). Ediciones Paidós.

Nota: *GEMINI-GOOGLE*. Este trabajo también ha contado con información y aportaciones obtenidas en conversaciones del autor, con la aplicación citada. Una vez validadas en fuentes externas las referencias, rechazados o admitidos autores, resúmenes y comentarios, a partir de las indagaciones e investigaciones efectuadas y con conocimientos y valor añadido propio del autor, este responde del texto ofrecido. *Ha recibido también aportaciones del colectivo ALGOVERIT y específicamente, agradece por su colaboración los comentarios de Ignacio Miguel González García y Bienvenido Pascual Encuentra.*