



ORGANIZA:



SDAQ2025

**SALÓN DEPARTAMENTAL DE ARTISTAS QUINDÍO 2025
DISRUPTIVO - NEXUS: IA / ARTE / HUMANO / NATURALEZA EN FLUJO**

DESCRIPCIÓN DEL TEMA CURATORIAL CON SUSTENTO TEÓRICO

AVANCES QUE (NOS) CAMBIAN LA VIDA

«Porque robando tu flor, el resplandor del fuego, origen de todas las artes, se la entregó a los hombres. Ha de pagar la pena a los dioses por una falta como esta». En esta obra de Esquilo, Prometeo es castigado por entregar «un honor a los mortales» que era solo para los dioses: el fuego.

<https://www.udep.edu.pe/hoy/2022/06/importancia-fuego/>

El fuego puede ser considerado el primer gran invento (algunos prefieren hablar de descubrimiento) de la humanidad, y su dominio representa un punto de inflexión en la evolución de los homínidos , ya que supuso una verdadera revolución demográfica al cambiar la forma de protegerse del frío, de alimentarse, de comunicarse. En su libro “Vigilar el fuego es vigilar la indefensión humana” (Marín, 2020), el autor plantea que gracias al fuego el hombre superó los obstáculos de los movimientos de rotación y de traslación. Fue el primer triunfo de la especie sobre los movimientos astrales con lo cual, venciendo la oscuridad y el frío de la noche y del invierno, haciendo posible el poblamiento de las zonas templadas y frías del planeta. Con el fuego se inauguró ese espacio que llamamos “hogar” donde hallamos resguardo para nuestra particular indefensión. Por todo ello y mucho más, el fuego resultó determinante en la evolución de nuestra especie. Metafórica y literalmente, por medio de este elemento, el hombre encontró la luz en un mundo de tinieblas.

Pues bien, todo parece indicar que nos encontramos en medio de un acontecimiento que puede asimilarse al descubrimiento y el control del fuego cuando dimensionamos su envergadura y su potencial evolutivo, a saber: el surgimiento de la Inteligencia Artificial (IA) a partir de los avances en el computador, la computación y la robótica. Y la inteligencia artificial como mediación de Vida Artificial. Súbitamente, se trastoca nuestra clásica visión antropocéntrica del mundo y estamos expuestos a tenérsela que ver formas de inteligencia y de vida nunca antes conocidas, nunca antes previstas. (Maldonado, 2024)

Por supuesto, estamos acostumbrados a que la tecnología, al ser una forma de relación social y con el mundo, siempre evoluciona. Pensemos tan solo en la rueda, la bombilla y el automóvil. Ninguna tecnología permanece fija. La tecnología empieza, se desarrolla, persiste, muta, se estanca y decae, igual que los organismos vivos El ciclo de vida evolutivo se produce en el uso y desarrollo de cualquier tecnología. Surge un nuevo núcleo de alta tecnología que desafía a las redes de apoyo tecnológico existentes, que se ven así obligadas a co-evolucionar con él. Se diseñan nuevas versiones del núcleo y se encajan en una red de apoyo tecnológico cada vez más apropiada, con efectos de alta tecnología cada

vez más pequeños. La alta tecnología se convierte en tecnología normal, con versiones más eficientes que encajan en la misma red de soporte. Por último, incluso las ganancias de eficiencia disminuyen, el énfasis se desplaza a los atributos terciarios del producto (apariencia, estilo), y la tecnología se convierte en tecnología apropiada que preserva la red de apoyo tecnológico. Este estado de equilibrio tecnológico se establece y fija, resistiéndose a ser interrumpido por una mutación tecnológica; entonces aparece una nueva alta tecnología y se repite el ciclo. (Milan Zeleny, 2009)

También hemos tenido experiencias con las tecnologías que pudieran catalogarse como “disruptivas”, esas que - según Clayton M. Christensen quien introdujo el concepto -, se destacan por ser innovaciones que crean nuevas formas de hacer las cosas, en primera instancia, en los escenarios del mercado o una industria existente al introducir nuevos productos, servicios o modelos de negocio. Pero estas tecnologías no solo alteran significativamente la forma en que las empresas operan y se conectan con los clientes, sino también la manera en que la sociedad funciona. Estas tecnologías pueden tener un impacto significativo en la forma en que vivimos, trabajamos y nos relacionamos con las máquinas y con el mundo. (Christensen, 1995).

Las tecnologías disruptivas impulsan la innovación y el progreso en diversos campos y pueden mejorar la calidad de vida de las personas a través de soluciones más eficientes y personalizadas. Pero es necesario tener en cuenta que la adopción de tecnologías disruptivas puede generar cambios significativos en las estructuras sociales y económicas, requiriendo procesos de adaptación y formación. El acceso a estas tecnologías puede ser desigual, lo que puede aumentar la brecha entre diferentes grupos de poblacionales. Las transformaciones en el universo del trabajo pueden generar desempleo en algunos sectores y requerir una adaptación en el mercado laboral. Al respecto, el Informe sobre el desarrollo mundial de 2019 del Banco Mundial sobre *La naturaleza cambiante del trabajo* examina cómo la tecnología da forma a la demanda relativa de determinadas competencias en los mercados laborales y amplía el alcance de las empresas: la robótica y las tecnologías digitales, por ejemplo, permiten a las empresas automatizar sus procesos, sustituyendo mano de obra por máquinas para ser más eficientes, e innovar, ampliando el número de tareas y productos.

Adicionalmente, los productos considerados innovaciones disruptivas tienden a saltarse etapas del proceso tradicional de diseño y desarrollo de productos para ganar rápidamente tracción en el mercado y ventaja competitiva. En uno de sus recientes boletines la Superintendencia de Industria y Comercio señala que ...”Colombia y el mundo han experimentado un creciente cambio económico y social hacia la transformación digital. En ese sentido, existen diversas tecnologías que, por ejemplo, brindan valor agregado a los procesos de comunicación e interacción desde el manejo y control de dispositivos en la

interacción humano-computador. Los desarrollos tecnológicos que el Centro de Información Tecnológica y Apoyo a la Gestión de la Propiedad Industrial (CIGEPI) identificó y analizó en esta área los agrupó en 2 tendencias: interacción humano-computador e interfaz humano-máquina.”...

<https://sedeelectronica.sic.gov.co/publicaciones/boletin-tecnologico/infraestructura/interaccion-humano-maquina>

De otra parte, a menudo se opone resistencia a la implantación de la alta tecnología. Esta resistencia es bien entendida por parte de los participantes activos en la red de apoyo tecnológico necesaria. El coche eléctrico será resistido por los operadores de gasolineras de la misma manera que los cajeros automáticos fueron resistidos por los cajeros de los bancos y los automóviles por los fabricantes de fustas. La tecnología no reestructura cualitativamente la red de apoyo tecnológico y, por tanto, no será resistida y nunca lo ha sido. Los mandos intermedios se resisten a la reingeniería de procesos de negocio porque ello representa un asalto directo a la red de apoyo (jerarquía coordinativa) de la que prosperan. El trabajo en equipo y la multifuncionalidad son resistidos por aquellos cuya red de apoyo tecnológico les proporciona la comodidad de una estrecha especialización y un trabajo dirigido por órdenes. (Zeleny, 2020)

Christensen, en su última publicación (2023) propone que la tecnología disruptiva también debe ser utilizada por el desarrollo en las naciones en vías de desarrollo para mejorar las condiciones económicas en las que las instituciones gubernamentales no trabajan. El concepto aparece en el texto coescrito con [Efosa Ojomo](#) llamado "la paradoja de la prosperidad".

En conclusión, puede afirmarse que la adopción de tecnologías disruptivas plantea cuestiones éticas y de responsabilidad social y ambiental, las cuales requieren convocar la participación de una amplia diversidad de actores, visiones e intereses. Pero, sin duda, el caso de la Inteligencia Artificial se revela como especialmente sensible, **cuando resulta importante que tomemos conciencia de que parte del futuro se va a diseñar con tecnología. Cómo caminamos hacia allí, cómo queremos que sea el futuro es importante y necesita de la participación de todos.** Y más allá, como advierte y argumenta Maldonado (2024), ...”expresamente, contra la idea generalizada por legos y expertos, **no se trata simple y llanamente de un fenómeno tecnológico.** Profundo motivos y preocupaciones se encuentran en su base. (...) **El centro de este trabajo pone en evidencia que la Inteligencia Artificial es en realidad Vida Artificial.”...**

REVISANDO NUESTRA COMPRENSIÓN DE “INTELIGENCIA” Y DE “INTELIGENCIA ARTIFICIAL”:

De acuerdo con Stuart Russell y Peter Norvig, autores del texto estándar sobre IA para estudiantes universitarios titulado *Inteligencia Artificial – un enfoque moderno*: ... Los hombres se han denominado a sí mismos como Homo sapiens (hombre sabio) porque nuestras capacidades mentales son muy importantes para nosotros. Durante miles de años, hemos tratado de entender cómo pensamos; es decir, entender cómo un simple puñado de materia puede percibir, entender, predecir y manipular un mundo mucho más grande y complicado que ella misma. El campo de la inteligencia artificial, o IA, va más allá: no sólo intenta comprender, sino que también se esfuerza en construir entidades inteligentes.”... La Inteligencia y la Vida Artificial constituyen proyectos que nos acercan a la comprensión de nosotros y de la vida misma, pero al mismo tiempo nos encaran al desafío de vernos coexistiendo con inteligencias que nos sobrepasan en muchos campos y con otras formas de vida que pudieran ser. En este sentido, se plantea incluso el reto de asistir a nuestra propia reinvención como especie...

El computador y la computación introducidos por A. Turing y Von Neuman son el antecedente de la IA y consisten en una articulación entre el componente material (hardware) y el componente de lógica o lenguajes de programación (software). ...”De lejos, la mejor herramienta jamás concebida es el computador o la computación, genéricamente llamada como una máquina para propósitos generales. Lo que esto quiere significar es con ella se puede hacer virtualmente cualquier cosa, literalmente. Producir y procesar imágenes, palabras, textos, música, construcciones de cualquier índole, acumular y guardar información, compartimentarla ilimitadamente, actuar a distancia con precisión total, ver las estrellas y los agujeros negros y los confines del universo, desentrañar la estructura de la materia y las escalas microscópicas antes inimaginables, y muchas otras.”...(Maldonado, 2024)

Como campo de investigación, la IA nace poco después de la Segunda Guerra Mundial en 1956 en Dartmouth College. El punto de partida propuesto por McCarthy, entonces un profesor de matemáticas, fue que “cada aspecto del aprendizaje y cualquier otra característica de la inteligencia puede ser descrita, en principio, con tal precisión, que una máquina podría simularla.”...(<https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth>). En contraparte, el interés de James Moor como filósofo se centraba en que el desarrollo cognitivo de las máquinas pudiera establecer un marco de comprensión mayor para la inteligencia humana. En todo caso, ambos énfasis han realizado un recorrido en paralelo; se trata del procesamiento de información.

En el momento de su concepción, Turing ya había planteado el problema de base: **¿pueden pensar las máquinas?** (Turing, 1950). Si bien el acercamiento tradicional a este problema comenzaría por definir los términos “máquina” e “inteligencia”, Turing decide tomar otro camino y reemplaza la pregunta inicial con una nueva, la cual está estrechamente relacionada y no se expresa mediante palabras ambiguas: **¿pueden las máquinas hacer, lo que nosotros (como entidades pensantes) hacemos?** Para evaluar la inteligencia de las máquinas y dar respuesta a esta nueva versión de la pregunta básica, Turing propone una prueba ampliamente conocida como “El test de Turing” consistente en: **¿es posible**

desarrollar una máquina de tal suerte que sea imposible distinguir si la interacción que tenemos con ella es con un ser humano o con la máquina? Dado que el tipo de interacción propuesto con el Test de Turing es conversacional, desde esta perspectiva, aunque es controvertida, la inteligencia se definió en términos de comportamiento lingüístico.

Turing, Alan (1952), «Can Automatic Calculating Machines be Said to Think?», en Copeland, B. Jack, ed., *The Essential Turing: The ideas that gave birth to the computer age*, Oxford: Oxford University Press, ISBN 0-19-825080-0.

(Mei, et al., 2024; Gonçalves, 2023; Geman et al., 2015).

Mei Q, Xie Y, Yuan W, Jackson MO. A Turing test of whether AI chatbots are behaviorally similar to humans. *Proc Natl Acad Sci U S A*. 2024 Feb 27;121(9):e2313925121. doi: 10.1073/pnas.2313925121. Epub 2024 Feb 22. PMID: 38386710; PMCID: PMC10907317.

En vista de que, a medida que la IA interactúa con los humanos en una gama cada vez mayor de tareas y teniendo en cuenta que gran parte de la programación de IA es patentada, es importante comprender cómo se comporta. Esto llevó a Mei et al (2024) a desarrollar un test de Turing para evaluar los rasgos de comportamiento y personalidad que exhibe la IA, específicamente en este caso variantes de ChatGPT. Los investigadores examinaron cómo se comportan los chatbots en un conjunto de juegos de comportamiento clásicos que están diseñados para provocar rasgos como la confianza, la equidad, la aversión al riesgo, la cooperación, etc., y evaluar cómo responden a una encuesta psicológica tradicional de Big5 que mide los rasgos de personalidad. Hallaron que ChatGPT-4 exhibe rasgos de comportamiento y personalidad que son estadísticamente indistinguibles de un humano al azar entre decenas de miles de sujetos humanos de más de 50 países. Los chatbots también modifican su comportamiento en función de la experiencia previa y los contextos "como si" estuvieran aprendiendo de las interacciones y cambian su comportamiento en respuesta a diferentes encuadres de la misma situación estratégica. Sus comportamientos son a menudo distintos de los comportamientos humanos promedio y modal, en cuyo caso tienden a comportarse en el extremo más altruista y cooperativo de la distribución. Los investigadores estimaron que actúan como si estuvieran maximizando un promedio de sus propios beneficios y los de su pareja.

La IA abarca en la actualidad una gran variedad de subcampos, que van desde áreas de propósito general, como el aprendizaje y la percepción, a otras más específicas como el ajedrez, la demostración de teoremas matemáticos, la escritura de poesía y el diagnóstico de enfermedades. La IA sintetiza y automatiza tareas intelectuales y es, por lo tanto, potencialmente relevante para cualquier ámbito de la actividad intelectual humana. En este sentido, es un campo genuinamente universal. (Russell y Norvick, 2008)

Si bien algunas veces se ha supuesto que los programas de IA vienen incluidos en algún robot con apariencia similar a la nuestra, la IA más bien se halla entre nosotros a la manera de una aspiradora automática que anda por toda la casa, de un software que nos escucha cuando encendemos el ordenador, del programa que detecta nuestra huella dactilar y que se encuentra instalado en nuestro teléfono, del televisor que nos recomienda qué ver o una

app que califica al ganador en un concurso de televisión. Tal vez durante la última semana utilizamos el Dalle – e o el ChatGPT para llevar a cabo nuestras tareas y actividades. En fin, hoy en día la IA ya nos acompaña en prácticamente cualquier área que podamos pensar; se trata de algo transversal, de algo que literalmente “nos atraviesa”, así no nos demos cuenta...Sin embargo, cuando escuchamos que un sistema de inteligencia artificial “cobró consciencia y siente”, también llegan las advertencias, los miedos, las dudas y las peticiones de regulación.

De acuerdo con Russell y Norvick, el principal tema unificador de este enorme campo es la idea del *agente inteligente*. Se define, entonces, la IA como el estudio de los agentes que reciben percepciones del entorno y llevan a cabo las acciones. Cada agente implementa una función la cual estructura las secuencias de las percepciones en acciones; también entran en consideración las diferentes formas de representar estas funciones. (Desde un punto de vista técnico éstas hacen referencia a sistemas de producción, agentes reactivos, planificadores condicionales en tiempo real, redes neurales y sistemas teóricos para las decisiones).

El aprendizaje juega un papel central. “La vieja IA” se basa en el procesamiento de símbolos y los conocimientos se adquieren de los expertos en un campo particular. Esta IA es mucho más transparente y opera, por decirlo así, “de arriba hacia abajo” (top down), y no es escalable. Esto significa que su introducción en entornos desconocidos limita el diseño del agente, favoreciendo así la representación y el razonamiento explícitos del conocimiento.

En contraste, la IA que tenemos hoy en día (“la nueva IA”) está basada en datos... Opera “de abajo hacia arriba” (bottom up) y deriva el conocimiento de su propio aprendizaje y de la extracción de patrones a partir de los datos con los cuales es alimentado el sistema. Esta nueva IA es capaz de generalizar y de predecir en muchas áreas, desde el lenguaje natural, la visión artificial o el aprendizaje automático. Un asunto ético y social de especial interés se pregunta por las fuentes de estos datos y la manera en que son obtenidos. De otra parte, la IA también requiere ingentes recursos energéticos y ambientales durante su operación.

Casi todos los sistemas de IA superan nuestra inteligencia; pero lo hacen en un campo determinado. Al día de hoy, el computador (en términos generales) ha pasado numerosas pruebas de Turing en nichos específicos como son el juego de ajedrez, el Go, el Mahjong, y ha mostrado ser capaz de llevar a cabo tareas y acciones, incluso tomar decisiones, superando la capacidad de los seres humanos.

La mayoría de las IA que tenemos son “de nicho estrecho”, esto es, capaces de exhibir un nivel de inteligencia muy alto en un campo muy concreto. Pueden ser estupendos diagnosticando un tipo de cáncer, pero no se desempeñan satisfactoriamente como médicos generales porque ello supone un conocimiento amplio. En cuanto a la consciencia y los sentimientos es preciso aclarar que ni siquiera sabemos cómo ocurren ciertos procesos en nosotros los seres humanos. Hay robots que pueden modelar sentimientos y puede parecer que cuentan con una consciencia real. Lo que hacen los chats es predecir la siguiente palabra en un texto, es decir que son un buscador muy sofisticado, pero no es capaz de razonar profundamente.

PRINCIPALES PREOCUPACIONES Y TEMORES QUE INSPIRA UNA “SUPERINTELIGENCIA”

Es difícil anticipar hasta dónde va a llegar la inteligencia artificial. Stephen Hawking argumentó que no existe una ley física que impida que las partículas se organicen de manera que realicen cálculos aún más avanzados que los arreglos de partículas en el cerebro humano; por lo tanto, la superinteligencia es físicamente posible. (Hawking, 2015). Además de las posibles mejoras algorítmicas en comparación con los cerebros humanos, un cerebro digital puede ser en varios órdenes de magnitud más grande y más rápido que un cerebro humano, cuyo tamaño se vio limitado por la evolución para ser lo suficientemente pequeño como para pasar por un canal de parto. (Graves, Matthew, 2015).

Y en medio de este escenario de apertura e incertidumbre, no faltan las visiones y posturas extremas que incluyen ese temor relacionado con *el riesgo existencial de la inteligencia artificial general*. Bajo el nombre de “riesgo existencial” se hace referencia a la hipótesis de que un progreso sustancial en la Inteligencia Artificial General (IAG) podría resultar en la extinción humana o alguna otra catástrofe global irreparable. Este temor se halla estrechamente vinculado al miedo a **perder el control y el dominio sobre la tecnología**. El razonamiento de base es el siguiente: ...”la especie humana actualmente domina a otras especies porque el cerebro humano tiene algunas capacidades distintivas de las que carecen otros animales. Si la IA supera a la humanidad en inteligencia general y se volviera ‘superinteligente’, entonces podría volverse difícil o imposible controlar para los humanos.”... Bostrom, Nick (2012).

A pesar de que la probabilidad de este tipo de escenario es ampliamente debatida y depende en parte de diferentes escenarios sobre el progreso futuro de la informática, destacados científicos informáticos y directores ejecutivos de tecnología como Geoffrey Hinton, el propio Alan Turing, Elon Musk y Sam Altman, director ejecutivo de OpenAI, han expresado su preocupación por la superinteligencia. La literatura, el cine y en especial el género de la ciencia ficción han producido ejemplares que muestran distintos escenarios posibles al respecto.

Resumiendo, entonces, existen tres fuentes de preocupación con relación a una Inteligencia Artificial General que pudiera tomar la forma de “superinteligencia”:

Ø La primera, como ya se dijo, tiene que ver con **perder el control y el dominio** sobre las máquinas producidas por el humano mismo.

En este contexto vale la pena traer a colación las Tres Leyes de la Robótica, junto con la Ley Cero, un legado de Isaac Asimov que sigue mencionándose cuando se trata de formular principios que guían el comportamiento de los robots y las tecnologías emergentes en su interacción con los humanos, sobre todo en contextos de la literatura y el cine. (Un ejemplo es la película de acción y ciencia ficción “Yo Robot” producida en 2004, dirigida por Alex Proyas y protagonizada por Will Smith)

Las Tres Leyes de la Robótica son:

- 1.- Un robot no puede dañar a un ser humano o, por inacción, permitir que un ser humano sufra daño.
- 2.- Un robot debe obedecer las órdenes dadas por los seres humanos, excepto si dichas órdenes entran en conflicto con la Primera Ley.
- 3.- Un robot debe proteger su propia existencia, siempre y cuando dicha protección no entre en conflicto con la Primera o Segunda Ley.

El temor por el control y el dominio se acompaña del miedo a “no poder apagar la máquina”. Sin embargo, por el contrario, escépticos como el informático teórico Yann LeCun argumentan que las máquinas superinteligentes no tendrán ningún deseo de autopreservación.

Ø Una segunda fuente de preocupaciones gira en torno a **la alineación de la IA**. En 2020, Brian Christian, un autor norteamericano de publicaciones no ficcionales, investigador, poeta y programador, presentó su último libro titulado *The Alignment Problem*.

Se considera que inculcarle a una máquina suprainteligente valores compatibles con los humanos puede ser un problema más difícil de lo que se supone ingenuamente. Dicho de otro modo, sería extremadamente difícil “alinearse” la superinteligencia con la amplitud completa de importantes valores humanos y aplicar restricciones.

La alineación de sistemas es parte de un campo de estudio más amplio llamado **seguridad en la inteligencia artificial** dedicado al estudio de cómo construir sistemas de inteligencia artificial que sean seguros. Las vías para la investigación de la alineación incluyen el aprendizaje de los valores y las preferencias humanas, el desarrollo de inteligencia artificial honesta, la supervisión extensible, el examen e interpretación de modelos de inteligencia artificial, y la prevención de comportamientos emergentes, como la búsqueda de poder.

La investigación de la alineación tiene conexiones con la investigación de la *interpretabilidad* (Rorvig, Mordechai, 2022); *la solidez* (Amodei, Dario et al, 2016 y Hendrychs, Dan et al, 2022); *la detección de anomalías*; *la incertidumbre calibrada*, (Rorvig Morderchai, 2022); *la verificación formal*, (Russell, Stuart et al, 2015); *el aprendizaje por preferencias* (Wirth, Christian et al, 2017; Christiano, Paul et al, 2017; Heaven, Will, 2022); *la ingeniería de seguridad* (Hendrychs, Dan et al, 2022); *la teoría de juegos* (Clifton, Jesse, 2020; Dafoe, Allan et al, 2021); *la equidad algorítmica* (Prunkl, Carina y Whittlestone, Jess, 2020); *las ciencias sociales* (Irving, Geoffrey y Askill, Amanda, 2029)

Ø La **singularidad tecnológica** constituye una tercera fuente de preocupación. Se trata de que el rápido desarrollo de las máquinas inteligentes pueda tomar por sorpresa a la especie humana. ...”Si imaginamos que una primera generación de un programa de

computación que puede igualar ampliamente la efectividad de un investigador de IA pudiera reescribir sus algoritmos y duplicar su velocidad o capacidades en seis meses, entonces cabría esperar que el programa de segunda generación tarde tan solo tres meses para realizar una parte similar del trabajo.

En este escenario, el tiempo para cada generación continúa reduciéndose, y el sistema experimenta una gran cantidad de etapas de mejora sin precedentes en un corto intervalo de tiempo, saltando de un desempeño infrahumano en muchas áreas a un desempeño sobrehumano en prácticamente todos los dominios de interés. "... (Good, I. y Vinding, M. 2026) Una verdadera "explosión de la inteligencia"...

Retomando a Russell y Norvick, pero esta vez en su escrito titulado "The Ethics and Risks of Developing Artificial Intelligence (2009), los temores y las preocupaciones se traducen en **tener en cuenta ciertas dificultades que son comunes tanto a los sistemas informáticos tipo IA como a los que no lo son**. Parten del hecho de que ..."casi cualquier tecnología tiene el potencial de causar daño en las manos equivocadas, pero con "una superinteligencia", tenemos el nuevo problema de que las manos equivocadas podrían pertenecer a la tecnología misma". Y concluyen: ..."Incluso si los diseñadores del sistema tienen buenas intenciones, es preciso tener en cuenta estas dificultades:

- La implementación del sistema puede contener errores inicialmente desapercibidos pero posteriormente catastróficos. Una analogía son las sondas espaciales: a pesar del conocimiento de que los errores en las costosas sondas espaciales son difíciles de corregir después del lanzamiento, históricamente los ingenieros no han podido evitar que ocurran errores catastróficos.
- No importa cuánto tiempo se dedique al diseño previo a la implementación, las especificaciones de un sistema a menudo dan como resultado **un comportamiento no deseado** la primera vez que se encuentra con un nuevo escenario. Por ejemplo, el bot Tay de Microsoft se comportó de manera inofensiva durante las pruebas previas a la implementación, pero se le provocó con demasiada facilidad un comportamiento ofensivo cuando interactuó con usuarios reales. (Dowd, Maureen, 2017)
- Los sistemas de IA, en particular, agregan un tercer problema, a saber: incluso con los requisitos "correctos", la implementación sin errores y el buen comportamiento inicial, las capacidades de aprendizaje dinámico de un sistema de IA pueden hacer que evolucione hacia un sistema con un comportamiento no deseado, incluso sin escenarios externos imprevistos. Una IA puede estropear en parte un intento de diseñar una nueva generación de sí misma y crear accidentalmente una IA sucesora que sea más poderosa que ella misma, pero que ya no mantenga los valores morales humanos compatibles preprogramados en la IA original. Para que una IA que se mejore a sí misma sea completamente segura, no solo debería estar libre de errores, sino que también debería poder

diseñar sistemas sucesores que también estén libres de errores.”...(Russell y Norvick, 2009; Yampolskiy, Roman V. 2014)

Nick Bostrom, filósofo sueco fundador del Future of Humanity Institute en la Universidad de Oxford, se atreve a ir aún más allá en los planteamientos en torno al *riesgo existencial*, al afirmar que estas tres dificultades se convierten en catástrofes en lugar de molestias en cualquier escenario en el que la superinteligencia etiquetada como "mal funcionamiento" predice correctamente que los humanos intentarán apagarlo y despliega con éxito su superinteligencia para burlar tales intentos: un escenario que ha recibido el nombre "giro traicionero" (treacherous turn.) Bostrom, Nick, 2015)

Por supuesto, una postura mucho más matizada surgió en el mismo año en que Bostrom publicó su libro *Superintelligence : paths, dangers, strategies*. Citando importantes avances en el campo de la IA y el potencial de que la IA tenga enormes beneficios o costos a largo plazo, la [Carta Abierta sobre Inteligencia Artificial](#) de 2015 declaró: ..."El progreso en la investigación de la IA hace que sea oportuno centrar la investigación no solo en hacer que la IA sea más capaz, sino también en maximizar el beneficio social de la IA. Tales consideraciones motivaron el Panel Presidencial AAAI 2008-09 sobre Futuros de IA a Largo Plazo y otros proyectos sobre impactos de IA, y constituyen una expansión significativa del campo de la IA en sí, que hasta ahora se ha centrado en gran medida en técnicas que son neutrales con respecto al objetivo. Recomendamos una investigación ampliada destinada a garantizar que los sistemas de IA cada vez más capaces sean *robustos y beneficiosos: nuestros sistemas de IA deben hacer lo que queremos que hagan*. Los firmantes incluyeron al presidente de AAAI, [Thomas Dietterich](#), [Eric Horvitz](#), [Bart Selman](#), [Francesca Rossi](#), [Yann LeCun](#) y los fundadores de [Vicarious](#) y [Google DeepMind](#).

Resulta claro que la IA no es solo un campo de investigación y exploración; también se ha convertido en una gran industria, atravesada por todo tipo de intereses. Algunos advierten que se intensifica un escenario propicio para el capitalismo: ..."El discurso transhumanista sobre la IA es teología política postmoderna. Una tecnocracia revestida, camuflada, de lenguaje tecnológico digital"...opina Francisco Garrido (2023) en su columna periodística.

“¿MÁS INTELIGENTE?”...

Fuera del campo de la inteligencia artificial, la "inteligencia" a menudo se usa de una manera que connota sabiduría moral o aceptación de formas "correctas" de razonamiento moral. En un extremo, si la moralidad es parte de la definición de inteligencia, entonces, por definición, una máquina superinteligente se comportaría moralmente. Sin embargo, la mayoría de las investigaciones de "inteligencia artificial" se centran en la creación de algoritmos que "optimicen", de forma empírica el logro de cualquier objetivo que los investigadores hayan especificado. (Bostrom, 2015)

Para evitar el antropomorfismo o el bagaje de la palabra "inteligencia", se puede pensar en una inteligencia artificial avanzada como un "proceso de optimización" impersonal que toma estrictamente las acciones que considera más probables para lograr sus objetivos (posiblemente complicados e implícitos). Otra forma de conceptualizar una inteligencia artificial avanzada es imaginar una máquina del tiempo que envía información hacia atrás

en el tiempo sobre qué elección que siempre conduce a la maximización de su función objetivo; esta elección se genera independientemente de cualquier preocupación ética externa. (Waser, Mark, 2011; Koebler, Jason, 2016)

¿Más inteligente significa, entonces, más seguro...o más peligroso...más poderoso?; ¿más autónomo?; ¿capacidad de emocionar e incluso amar?; ¿adquirir conciencia?; ¿estar en la capacidad de realizar cualquier tarea humana y reemplazar toda la fuerza laboral?; ¿resolver problemas complejos que nosotros no podemos resolver?; ¿el impulso a la autopreservación?; ¿el reconocimiento de sus propios derechos? (como se plantea respecto a la rebelión de los robots); ¿comenzar a mostrar rasgos humanos, tales como una moralidad?; ¿adquirir características espirituales?; entre otras concepciones...

En todo caso, lo que se entiende por inteligencia tiene que ver estrechamente con

RESOLVER PROBLEMAS de distinto tipo y complejidad.

Es posible advertir que la comprensión de la inteligencia puede tomar distintos caminos y que de ello depende lo que entendamos bajo el término "Inteligencia Artificial". Russell y Norvick presentan en su libro una síntesis de distintas definiciones y concepciones:

ü Algunas se refieren a **procesos mentales y al razonamiento**, mientras que otras se enfocan **en la conducta**.

ü Algunas miden el éxito **en términos de la fidelidad en la forma de actuar de los humanos** (recordándonos al Test de Turing), y otras toman como referencia **un concepto ideal de inteligencia que los autores llaman "racionalidad"**. Un sistema es racional si hace "lo correcto", en función de su conocimiento.

Los autores agregan que el surgimiento de conflictos y enfrentamientos entre estos enfoques resulta previsible, dado que unos se centran en el comportamiento humano y por lo tanto se alimentan de una ciencia empírica (incluye hipótesis y confirmaciones mediante experimentos), mientras que el enfoque racional implica una combinación de matemáticas e ingeniería. Sin embargo, resulta interesante que los diversos grupos, al mismo tiempo, se han ignorado y catalizado mutuamente.

Por mencionar tan solo algunos, en el campo interdisciplinar de la IA han sido relevantes los aportes desde:

- v la filosofía: ¿se pueden utilizar reglas formales para extraer conclusiones válidas?; ¿cómo se genera la inteligencia mental a partir de un cerebro físico?; ¿de dónde viene el conocimiento?; ¿cómo se pasa del conocimiento a la acción?
- v las matemáticas: ¿qué reglas formales son las adecuadas para obtener conclusiones válidas?; ¿qué se puede computar?; ¿cómo razonamos con información incierta? lógica, computación y probabilidades
- v la ingeniería computacional: ¿cómo se puede construir un computador eficiente? Para que la inteligencia artificial pueda llegar a ser una realidad se necesitan

dos cosas: inteligencia y un artefacto. El computador ha sido el artefacto elegido.

- v la economía: ¿cómo se debe llevar a cabo el proceso de toma de decisiones para maximizar el rendimiento?; ¿cómo se deben llevar a cabo acciones cuando otros no colaboren?; ¿cómo se deben llevar a cabo acciones cuando los resultados se obtienen en un futuro lejano? Estas cuestiones han sido exploradas por la teoría de la decisión y la teoría de juegos.
- v las neurociencias: ¿cómo procesa información el cerebro? Y más recientemente, ¿cómo procesa información todo el cuerpo?
- v la psicología cognitiva: ¿cómo piensan y actúan los humanos y los animales?; ¿cómo aprenden?
- v Teoría de control y cibernética (desde el año 1948 hasta el presente): ¿cómo pueden los artefactos operar bajo su propio control?

La teoría de control moderna, especialmente la rama conocida como control óptimo estocástico, tiene como objetivo el diseño de sistemas que maximizan una función objetivo en el tiempo. Lo cual se asemeja ligeramente a nuestra visión de lo que es la IA: diseño de sistemas que se comportan de forma óptima.

- v Lingüística: ¿Cómo está relacionado el lenguaje con el pensamiento?

La lingüística moderna y la IA «nacieron», al mismo tiempo y maduraron juntas, solapándose en un campo híbrido llamado lingüística computacional o procesamiento del lenguaje natural.

El problema del entendimiento del lenguaje se mostró pronto mucho más complejo de lo que se había pensado en 1957. El entendimiento del lenguaje requiere la comprensión de la materia bajo estudio y de su contexto, y no solamente el entendimiento de la estructura de las sentencias. Lo cual puede parecer obvio, pero no lo fue para la mayoría de la comunidad investigadora hasta los años 60.

ANALOGÍAS Y METÁFORAS QUE FUNDAN MITOS PROBLEMÁTICOS

El enfoque computacional del modelo cognitivo y la analogía entre el cerebro y el computador

Considerar que la inteligencia está íntimamente ligada al “pensar como un humano” nos remite, al menos en primera instancia, al *enfoque computacional del modelo cognitivo*. Ahora bien, para poder afirmar que un programa cualquiera piensa como un humano es preciso conocer los mecanismos que subyacen a ese modo particular de pensar. Se ha supuesto que al contar con una teoría suficientemente precisa acerca de cómo trabaja la mente humana, sería posible expresarla como un programa de computador. Y si los datos de entrada/salida del programa, así como los tiempos de reacción son comparables con los de un humano, podría confirmarse su similitud. Se reconocen aquí las huellas del Test de

Turing. Su propuesta fue retomada por investigadores como Allen Newell y Herbert Simon, quienes desarrollaron el así denominado “Sistema de Resolución General de Problemas” en 1961. Su interés no se limitó a comprobar que el programa resolviera correctamente los problemas propuestos, sino comparar etapas del procesamiento de información realizado por la máquina y por un humano encarado al mismo problema.

En este contexto resulta importante hacer referencia a un término clave en el campo de la IA, a saber: el *algoritmo*.^[1] Al respecto se maneja una analogía que se ha convertido en un mito extendido; el algoritmo como “receta de cocina”. Esto es, un algoritmo como secuencias de pasos con reglas muy definidas. Tendríamos, entonces un problema, una lista de pasos, programamos esos pasos, los realizamos y obtenemos una salida. Hoy en día esa definición resulta incompleta y problemática. Su comprensión se ha expandido, ha evolucionado. Así, con el tema de la *computación distribuida*, el procesamiento y los algoritmos operan en una red; se rompe el tiempo y la idea de orden y pasos secuenciales ya no es tan clara; se requieren otros mecanismos de coordinación.

También aparece el tema de los *algoritmos aleatorios* que hacen referencia a que dentro de sus comportamientos suceden decisiones aleatorias que permiten seguir avanzando en mejores espacios de búsqueda (optimización) de soluciones. Y también están los *algoritmos probabilísticos*. Ya no hay reglas definidas, tan solo ciertos márgenes donde se pueden mover los algoritmos. En todo caso, no es ya una receta de cocina, a no ser que entendamos el cocinar como algo caótico. En resumen, hay una cantidad de formas de procesamiento de información que resuelven problemas, precisamente porque no son secuenciales, porque están distribuidos y están sujetos a cosas externas que no pueden controlar como el tiempo, la aleatoriedad y la oscuridad (no sabemos siquiera como opera un algoritmo).

Las analogías y metáforas en torno al algoritmo están íntimamente ligadas a **la comprensión del cerebro como un computador** y a la visión más general de **los sistemas vivos como máquinas**, específicamente en este caso con relación al procesamiento de información.

Una analogía es un atajo mental para poder comprender un concepto. Inicialmente, se revela como un apoyo, una especie de bastón que nos ayuda en la aproximación a un nuevo campo de exploración. Pero las analogías tienden a volverse automáticas y cuando resulta decisivo ampliar la complejidad de lo que buscamos comprender, se convierten en un obstáculo. Culturalmente, es difícil replantearse algunas analogías y metáforas que se han vuelto costumbre. (...“cazarte fuertemente con una analogía puede llevarte a un bastón intelectual que no se puede dejar después.”...Chacón, C. 2024).

<https://www.youtube.com/watch?v=MVBV-GSm7Ms&list=TLPQMjUwNDlwMjWwpi9yScHwQ&index=2>

Como ya se mencionó, estrechamente relacionada con la analogía entre algoritmos y recetas de cocina, se halla la analogía entre el cerebro y el ordenador que tanto éxito ha tenido en las ciencias de la computación en la segunda mitad del siglo XX: un sistema que recibe, almacena y procesa información. En este caso, las analogías y metáforas no solo limitan la comprensión, sino que resultan sensiblemente problemáticas, desviando la atención hacia caminos sin salida o caminos no deseables.

Se supone que todas nuestras actividades cerebrales son consecuencia de la sucesiva acción de una serie de reglas (semejantes a las instrucciones de un programa informático) sobre contenidos informativos (representaciones), contenidos equivalentes a los datos que procesa el ordenador. La tarea de la psicología sería descubrir las reglas o procedimientos que en cada caso están a la base de nuestra cognición: la percepción, la memoria, el razonamiento... En cierto modo como si se buscara el *auténtico software de la mente*. Esta metáfora ha sido muy utilizada por la [psicología cognitiva](#) como guía en sus investigaciones sobre la mente y la conducta, y supone una concepción extremadamente racionalista del psiquismo humano, acercando este enfoque al campo tradicional de la IA.

De acuerdo con la teoría computacional de la mente: ...”La ‘mente’ ya no ‘se vuelve más fantasmal que un fantasma’. En cambio, la mentalidad enferma puede ser entendida sin pérdida de alcance o rigor, en los términos científicos de la neurofisiología ... y la biofísica matemática, la teoría contribuye con una herramienta para el tratamiento simbólico riguroso de las redes (neuronales) conocidas y un método fácil de construir redes hipotéticas de propiedades requeridas.(McCulloch & Pitts, 1943, p. 20-21)[2]

Como advertiera Karl Lashley en 1951, confiar en metáforas basadas en máquinas corre el riesgo de simplificar en exceso un órgano profundamente intrincado y evolucionado. Así como generaciones anteriores utilizaron metáforas hidráulicas o telefónicas para explicar el funcionamiento del cerebro, la analogía computacional también podría estar limitando nuestra comprensión.

El cerebro humano, a menudo descrito como el objeto más complejo del universo, sigue desconcertando a los científicos a pesar de las vastas investigaciones y avances tecnológicos dedicados a comprenderlo. La neurociencia moderna ha logrado capacidades asombrosas, como permitir que personas paralizadas controlen extremidades robóticas con su mente o incluso manipular la memoria y la percepción en animales y humanos. Sin embargo, a pesar de estos avances, existe una profunda inquietud en la comunidad científica acerca del marco teórico—o la ausencia del mismo—que guía esta investigación.

Muchos neurocientíficos argumentan que estamos llegando a un punto muerto en la comprensión del cerebro. A pesar de la avalancha de datos generados por iniciativas de investigación masivas, nuestra comprensión de cómo funciona el cerebro en su totalidad parece más fragmentada que nunca. Esta paradoja se resume de manera contundente en las palabras del neurocientífico francés Yves Frégnac: "**El big data no es conocimiento**". La acumulación de hechos, aunque impresionante, a menudo abre nuevas capas de complejidad, haciendo que la comprensión

global sea cada vez más esquivada. Si bien los avances de la neurología y la neurociencia ya permiten, gracias a la [resonancia magnética funcional \(fMRI\)](#), dibujar una suerte de mapa del cerebro, no acabamos de entender del todo cómo funciona. *El cerebro sigue siendo un misterio...*

El neurocientífico francés Romain Brette ha desafiado uno de los aspectos más omnipresentes de esta metáfora: la idea de la "codificación" neuronal. Aunque la codificación sugiere una representación de información que puede descifrarse, esta suposición implica la existencia de un observador ideal dentro del cerebro que descifre estas señales. En realidad, cómo las redes interconectadas del cerebro procesan y actúan sobre estas señales sigue siendo un misterio en gran medida. La crítica de Brette destaca el peligro de confundir una metáfora con la realidad.

El cerebro no es un receptor pasivo de estímulos. No se limita a absorber y almacenar datos como lo haría un disco duro de computadora. Por el contrario, como argumenta el neurocientífico húngaro György Buzsáki en *The Brain from Inside Out*, *el cerebro construye activamente información. Constantemente prueba posibilidades, busca soluciones e interactúa con el entorno de maneras que trascienden el modelo simplista de entrada-salida propio de la computación. Este juego dinámico resalta el papel del cerebro como un órgano moldeado por la evolución, diseñado para la acción y la supervivencia, no para procesar datos de manera aislada.*

De entrada hay que señalar que la metáfora nace ya lastrada por el dualismo entre mente y cerebro, y entre individuo y entorno. La cognición humana trasciende las fronteras del órgano específico; es una propiedad emergente del organismo, sus actividades y sus entornos, no una cualidad específica de ningún órgano. El sujeto de la cognición humana no es el cerebro sino la actividad mental, el metabolismo social y ecológico; es decir, la interacción activa entre la especie y el entorno.

¿Puede una computadora reproducir la enorme complejidad del metabolismo social y ecológico? De momento está muy lejos de ello y ni siquiera este ambicioso objetivo se encuentra entre sus metas programáticas originales. Establecer como objetivo ingenieril la mimesis artificial de la cognición del metabolismo social y ecológico humano es una meta imposible de la misma naturaleza que alcanzar en un momento dado la línea del horizonte: la aproximación al horizonte es siempre asintótica. Una IA idéntica a la cognición del metabolismo social y ecológico ya no sería idéntica, porque la misma IA habría ampliado la complejidad de dicho metabolismo... Si repasamos la historia de la computación veremos que ni Cantor, ni Hilbert, ni Gödel, ni siquiera Turing tenían en su programa de investigación ese objetivo.

Erik Larson, un científico e investigador pionero en el procesamiento del lenguaje natural, además de empresario tecnológico que trabaja a la vanguardia de la IA, presenta en su libro *El Mito de la Inteligencia Artificial*, cinco mitos sobre los discursos actuales de la IA: (a) La absoluta novedad de la misma, (b) la irreversibilidad de su desarrollo actual, (c) la equivalencia con la cognición humana, (d) La neutralidad ideológica y axiológica de la IA y

(e) la inminente superación de cualquier forma de cognición natural a cargo de la Inteligencia Artificial General. (Larson, E. 2022)

Como observó Francis Crick, el cerebro es una estructura compuesta, con diferentes regiones que evolucionaron para funciones específicas. Los neurocientíficos Anne Churchland y Larry Abbott han sugerido que nuestra comprensión final del cerebro podría parecerse a un mosaico: una colección de explicaciones altamente específicas unidas entre sí, más que una teoría unificada. Este enfoque fragmentado puede ser la única forma viable de captar un órgano tan diverso en su estructura y función.

Thomas Parson en su libro *Cyberpsychology and the Brain: The Interaction of Neuroscience and Affective Computing*, afirmaba que el cerebro no es un computador (una máquina) de propósitos generales con un procesador central unificado (una CPU). En contraste, la actividad funcional cerebral se comprende mucho mejor *en términos de una red neuronal* de gran escala que representa distintos subsistemas y relaciones entre ellas: lenguaje, reconocimiento facial, atención espacial, memoria, emoción, comportamiento ejecutivo.

Aunque las redes neuronales artificiales se inspiran en la estructura neuronal del cerebro, difícilmente se puede afirmar que emulan realmente su funcionamiento. Sin embargo, es en el campo de las redes neuronales artificiales, especialmente en las arquitecturas que se emplean en “deep learning, donde la IA está obteniendo algunos de los mayores avances.

La dependencia de metáforas como la computadora refleja un problema más amplio en el cientificismo mecanicista: la suposición de que los misterios del cerebro pueden explicarse completamente a través de datos empíricos y analogías tecnológicas. Esta visión, aunque productiva para generar hipótesis y experimentos, corre el riesgo de pasar por alto la naturaleza activa, encarnada y evolutiva del cerebro. Al aferrarnos a metáforas basadas en tecnologías contemporáneas, podríamos estar limitando nuestra imaginación y el alcance de futuros descubrimientos.

A medida que cuestionamos los límites de la metáfora computacional, el desafío radica en encontrar nuevas formas de conceptualizar el cerebro. Lo que está claro, sin embargo, es que comprender el cerebro requiere superar las restricciones del pensamiento mecanicista. El cerebro no es una computadora. Es un órgano de acción, percepción y construcción, un sistema dinámico que desafía las analogías simples. Reconocer los límites de nuestros marcos actuales es el primer paso hacia una comprensión más profunda y holística de este órgano extraordinario.

Nuevas comprensiones en el campo de la IA:

Alejandra Ciria (psicóloga de la UNAM) y Bruno Lara (responsable del laboratorio de Robótica Cognitiva en el Estado de Morelos) presentan una breve síntesis de algunas comprensiones más recientes en el campo de la IA:

<https://www.cienciascognitivascorporizadas.com/historia/met%C3%A1fora-de-la-mente>

- v La manipulación de símbolos físicos, noción central para la creación de máquinas inteligentes, de acuerdo a la Teoría computacional de la mente, y que según Newen y Simon (1976), tenía la capacidad necesaria y suficiente para realizar acciones inteligentes apropiadas para el sistema y adaptarse a las demandas del entorno, cede ante la **teoría del conexionismo**, un primer intento de superar la idea de que el cerebro se debe entender como un sistema lógico de manipulación de símbolos.

Para los conexionistas, el cerebro debía estudiarse y entenderse desde su neurofisiología, sus conexiones neuronales y su arquitectura basada en redes neuronales jerárquicas. Dentro del conexionismo, aunque las representaciones mentales ya no se entienden como la mera manipulación de símbolos mediante un conjunto de reglas, continúan siendo necesarias para explicar a la cognición como la actividad de resolver problemas mediante el procesamiento de información.

- v En 1980 John Searle propone otro experimento mental que denominó “la habitación china”. Lo que quería mostrar es que aunque un programa informático o una computadora, sea capaz de pasar la prueba de Turing (Turing, 1950), en realidad *no comprende el significado de ninguno de los símbolos que manipula*. Lo que está haciendo el programa es manejar de forma correcta las reglas de sintaxis, sin embargo no conoce la semántica de las oraciones. Una condición necesaria para la semántica es una relación casual apropiada de las entradas y salidas de una computadora - o un programa - con el resto del mundo, por lo que **la manipulación de símbolos (sintaxis) no es suficiente para tener semántica** (Searle, 1990).

Sin lugar a dudas, un *sistema informático experto* que juega al ajedrez o responde a preguntas naturales complejas muy rápidamente, o que incluso pasa la prueba de Turing, demuestra un alto dominio de experiencia en una tarea muy específica. Sin embargo, este sistema experto se limita a lo que ha sido diseñado y, lo que es más importante, *no puede comprender ninguno de los procesos subyacentes a su aparente comportamiento inteligente*. Por lo tanto, algunos de los investigadores de IA que se enfocaron en construir máquinas que actúan y piensan como lo hacen los humanos, en lugar de sólo diseñar sistemas expertos, se interesaron en **la hipótesis de la cimentación física para la adquisición de conocimiento**.

- v En 1989 Harnad sugirió que el significado de los símbolos se basa en categorías perceptuo-motoras. Específicamente, Harnad propuso un sistema que usa las proyecciones sensoriales de objetos y eventos distales para discriminar representaciones icónicas de modo que el sistema simbólico identifique las características invariantes de la entrada, dando una representación de categoría como salida. La *adquisición de significado* de los símbolos requiere de la identificación del *referente desde distintas perspectivas en el mundo*. Esto es lo que se conoce como **el problema de la cimentación de símbolos**, esto es, cómo la adquisición del significado del símbolo se basa en el mundo físico (Harnad, 1990).

Al mismo tiempo, Brooks (1990) intentó resolver el problema de la cimentación de símbolos mediante implementaciones en agentes artificiales autónomos. Sus propuestas y observaciones han tenido efectos de largo plazo en la comunidad de IA en general, particularmente, en el área de robótica. Una de sus aportaciones más importantes fue proponer que *el mundo es el mejor modelo de sí mismo*, por lo que un agente puede actuar utilizando la información que percibe y no requiere adquirir un modelo del mundo (Brooks, 1991). Por lo tanto, cuando el significado del símbolo se basa en el mundo físico, la necesidad de representaciones simbólicas tradicionales se desvanece por completo.

- v Las anteriores comprensiones han llevado a concluir que ***se requiere de un cuerpo que interactúe con el mundo para poder llevar un comportamiento inteligente.***

La cimentación física de la cognición implica tener un cuerpo con el cual se pueda interactuar con el entorno de tal forma que se puede procesar la información para otorgarle un significado situado en un contexto (Barsalou, 2008). Estas ideas llevaron a un cambio en el paradigma de la inteligencia artificial en el cual *tener un cuerpo es un prerequisite para la inteligencia y es necesario para la cognición* (Pfeifer y Bongard, 2006).

UN PRIMER GRAN SALTO: SUPERAR LA METÁFORA COMPUTACIONAL ALGORÍTMICA Y EL ENCEFALOCENTRISMO

Pensar a la inteligencia como corporizada (corporeizada) tiene importantes implicaciones para la comprensión de la cognición tanto biológica como artificial. Esto involucra considerar que los ***procesos cognitivos no son independientes del cuerpo***, sino que están fuertemente limitados y, al mismo tiempo, son posibilitados por éste. En este sentido, las capacidades cognitivas de los agentes biológicos y artificiales tienen su base esencial en la corporalidad, en su morfología particular y las propiedades materiales de sus cuerpos (Pfeifer y Bongard, 2006).

Desde esta nueva perspectiva, se considera que para comprender el funcionamiento del cerebro se debe comprender cómo es que éste se encuentra embebido en el sistema físico y cómo es que los agentes interactúan con su entorno. Una de las implicaciones más profundas de esta nueva forma de entender la cognición se centra en la relación entre procesos físicos y aquellos relacionados con el procesamiento de la información neuronal y de control (Hoffmann y Pfeifer, 2012). Por lo tanto, la ***concepción corporizada (corporeizada) de la inteligencia y de la cognición*** enfatiza la *conexión entre cuerpo, cerebro y entorno*.

¿QUÉ IMPLICA, ENTONCES, ACTUAR DE FORMA RACIONAL?

El mencionado salto constituye una cuestión que está a la base del enfoque del agente racional, entendiéndolo como algo que razona. Y entonces de los agentes informáticos

entendidos como agentes racionales se espera algo más que llevar a cabo inferencias correctas, como lo haría un programa convencional. Deberían estar dotados de controles autónomos; lograr percibir el entorno y persistir durante un período de tiempo prolongado; adaptarse a los cambios y estar en capacidad de alcanzar objetivos diversos. En fin, ...un agente racional es aquel que actúa con la intención de alcanzar el mejor resultado o, cuando hay incertidumbre, el mejor resultado esperado. "... (Russell y Norvick, 2008) En la vida hay situaciones para las que no hay nada correcto que hacer y en las que de todos modos es preciso tomar una decisión. Y existen, también, formas de actuar racionalmente que no implican realizar inferencias.

Desenvolverse en una sociedad y, en general, en ambientes complejos requiere la capacidad de generar sentencias comprensibles en lenguaje natural. Para aprender no basta la erudición; es necesario profundizar en el conocimiento de *cómo funciona el mundo para desarrollar las mejores estrategias de moverse y comportarse en él*. Los órganos de los sentidos apoyan la obtención de una idea más elaborada acerca de lo que una acción puede llegar a representar. Y si no se encuentra una solución, es preciso que el programa nunca pare de buscarla.

Al decidirse a favor de un **enfoque del diseño de un agente racional** y no limitarse al cumplimiento de "las leyes del pensamiento", se obtienen ventajas significativas. Como se expuso arriba, se trata de un enfoque mucho más amplio y no antropocéntrico. Podría decirse, que el comportamiento humano se adapta bien a entornos y situaciones específicas, y que es producto de un proceso evolutivo complejo, en gran medida desconocido y lejos de haber concluido...

Los intentos de desarrollar agentes racionales artificiales nos enseñan mucho acerca del procesamiento de información. Así, por ejemplo, develan que en entornos complejos no es posible obtener una racionalidad perfecta entendida como hacer siempre lo correcto. La demanda computacional que esto implica es demasiado grande. Por ello es mejor adoptar el punto de partida de una racionalidad limitada, entendida como actuar adecuadamente cuando no se cuenta con el tiempo suficiente para efectuar todos los cálculos que serían deseables.

LOS SISTEMAS VIVOS NO SON MÁQUINAS...

Hemos arribado así a la revisión de otra analogía o metáfora que es preciso superar: **Los sistemas vivos no son máquinas, en ningún sentido**. Ciertamente son sistemas físicos, pero no se reducen, en manera alguna, a la física; y manifiestamente no a la física clásica.

En la perspectiva de esta contundente afirmación se revisa el concepto de *hipercomputación biológica*, introducido por Stepney, 2009 y Maldonado, Gómez, 2015). Como tal, el término "hipercomputación" surgió a finales del siglo pasado haciendo referencia a que la computación que se conocía era finita y que cuando un problema tiene a ser infinito se consideraría un problema no computable, es decir, fuera del ámbito de la máquina de Turing y la Tesis de Church Turing, paradigmas definitorios en ese momento. Una manera de computar lo no computable (llevar a cabo super – tareas) se denominó hipercomputable. (Copeland, Jack, 2002). Sin embargo, esta búsqueda la cual apeló a

recursos en la teoría cuántica y en la teoría de la relatividad, configuró tan solo la hypercomputación clásica. (Davies y Gregersen, 2010).

Ahora bien, lo que interesa aquí es acercarnos a una mayor comprensión *del modo en que los sistemas vivos procesan información*. Revisemos algunos rasgos claves que resaltan Maldonado y Cruz en esta otra búsqueda centrada en encontrar nuevas formas de resolver problemas y computar cosas (Calude y Paun, 2001). Pasamos de la hypercomputación clásica la hypercomputación biológica no-clásica, superando los límites de lo que habitualmente se ha comprendido por “computar”:

- v Nos alejamos, como ya vimos, de esa expresión cultural y científica de la CPU que se condensa en el encefalocentrismo y, entonces, de la creencia de que el cerebro es la unidad determinante de los sistemas vivos, y específicamente de ciertos sistemas vivos. En cambio,

la hypercomputación biológica (HB) nos acerca a la explicación de los procesos computacionales en los sistemas biológicos, desde los genes a las bacterias, desde la célula hasta la biosfera. En este sentido, permite comprender lo que los sistemas vivos hacen para ser tales (Mitchell, 2012), abarcando tanto la esfera del desarrollo como de la evolución de aquellos. (Ignatova et. al., 2008).

- v Nos alejamos, también, de considerar que la “inteligencia” consiste simplemente en inteligencia o complejidad algorítmica. *Los sistemas vivos no son algorítmicos* (Syropoulos, 2008). Cuando consideramos que un sistema vivo es algorítmico, lo reducimos a un simple objeto físico clásico, susceptibles de ser explicado, por ejemplo, mediante mecanismos de acción-reacción (Ball, 2004). Los sistemas vivos carecen de cualquier solución algorítmica. Lo contrario sería tanto determinismo como reduccionismo, y ambos le prestan un flaco favor a comprender la complejidad de la vida, esto es, de los sistemas vivos.

- v Los sistemas de reparación metabólicos constituyen el mejor ejemplo de que los sistemas vivos no son máquinas en ningún sentido y no pueden ser simulables o computables en términos de una Máquina de Turing o de la teoría computacional Church - Turing.

El sistema inmunológico constituye un caso evidente de hypercomputación biológica interactiva y emergente. (Goldin et. al., 2006)

Tampoco encajan en las visiones clásicas *las redes y sistemas que hacen que los sistemas vivos sean sistemas de complejidad creciente*.

En las bacterias, el hardware varía en la misma medida en que la información se acumula y se procesa. Así, las bacterias pueden intercambiar información entre sí en la forma de hardware, por ejemplo como material genético (Ben-Jacob, 2009). En

general, los sistemas vivos no solo computan mediante números, sino mediante estructuras; apelan a las formas, a la geometría.

v Dado que los sistemas vivos no poseen toda la información que requieren para seguir vivos, es preciso que estén en capacidad de *sintetizar información relevante, de cara a problema relevantes*. (Ehrenfeucht et al., 2012) La computación de los sistemas vivos es emergente y autoorganizada y el procesamiento de información ocurre en más de una sola escala, de manera no secuencial y en dependencia de un operador externo.

Que sinteticen información significa que no solamente exploran el medioambiente permanentemente, sino que además y fundamentalmente, producen nueva información y la introducen en el entorno. *Los sistemas vivos son permanentes generadores de nueva información en el medioambiente*. (Maldonado y Cruz, 2015)

v El hecho de que los sistemas vivos sean esencialmente abiertos e incompletos, los encara al reto de buscar constantemente en el entorno materia, energía e información que los haga viables...(con todo y el riesgo de que en ocasiones esto no se logre). La distinción entre los sistemas vivos y el medioambiente es meramente analítica o epistemológica; en realidad no puede trazarse una clara línea demarcatoria entre ambos. Como ya se mencionó arriba, los sistemas vivos evolucionan al unísono con el propio medioambiente.

Adicionalmente, el medioambiente es intrínsecamente no-algorítmico y no-computacional. Dicho de otro modo, “el medioambiente” es un concepto esencialmente indeterminado. Allí el azar desempeña un papel activo, por lo cual no puede ser descartado en manera alguna (Kari y Rozenberg, 2008).

Los sistemas vivos computan y crean siempre nuevas funciones que difícilmente pueden ser entendidas ex ante.

...”En otras palabras, una Máquina de Turing es tan sólo un caso particular de un fenómeno mucho más general que es la Hypercomputación Biológica. Dicho en términos más explícitos, los algoritmos son tan sólo un caso particular de razonamientos y procesamientos que incluyen además (¡y fundamentalmente!) incertidumbre, aprendizaje, adaptación, interacción y sorpresa, y creatividad (incluyendo lógicas no clásicas.) En la naturaleza, los sistemas vivos procesan mucho más que números y funciones, lo cual significa exactamente que los sistemas vivos crean mundo y naturaleza. (...)

La mente –no el cerebro- es un sistema de respuesta de los sistemas vivos al entorno. *Pero la forma como los sistemas vivos reaccionan al medioambiente –esto es, literalmente, a la física- es creando nuevas realidades e introduciendo nuevas posibilidades que antes no*

existían. Esto define a un sistema vivo exitoso (fittest) y en esto exactamente consiste la adaptación. Dicho inversamente: cuantas menos posibilidades tenga un sistema vivo de reaccionar al entorno (= crear nuevos mundos) tanto más expuesto estará y se puede convertir en un organismo o especie que pudiera desaparecer.”... (Maldonado, 2016)

Hemos alcanzado así una comprensión mucho más compleja de “inteligencia” en el caso de los sistemas vivos. La inteligencia se revela en la constante exposición al problema agónico de seguir existiendo, de permanecer vivo, de continuar aprendiendo. Y, entonces, la inteligencia deviene “fitness”, adaptabilidad creativa e invención.

CUANDO EL ARTE Y LA VIDA SE TRASLAPAN EN ESPACIOS DE CREATIVIDAD

Vemos de qué modo el arte y la vida no solo se acercan, sino que se traslapan, una vez que procesos fundamentales tales como el aprendizaje, la adaptación, la autoorganización y la comunicación no se pueden explicar mediante una aproximación algorítmica normal. En términos culturales y sociales lo algorítmico adquiere la forma de normas, mandamientos, preceptos, recetas, planes y estrategias. Sin embargo, la comprensión de los sistemas vivos no se puede reducir a éstas expresiones y el horizonte que se abre ante nosotros es el de la creatividad, la innovación, el juego, los desafíos y las apuestas, la contingencia y la adaptación. ***Para los sistemas vivos computar significa vivir, (Maldonado, 2016) y vivir significa crear, en todos las esferas y ámbitos.***

La hypercomputación biológica nos permite incluso dar un paso más: ***el arte como vida y la vida como arte*** nos invita a superar un enfoque antropocéntrico, antropológico y antropomórfico de los sistemas vivos; nos introduce en la forma como los sistemas vivos procesan información conjuntamente, a la manera de una red de redes entrelazadas, ***una verdadera trama de la vida.*** (Kampis, 1996; Tsuda et al., 2006).

Maldonado (2016)

Revista Simbiótica vol. 3, n. 1, jan.-jun., 2016

Revista Simbiótica - Universidade Federal do Espírito Santo - Núcleo de Estudos e Pesquisas Indiciárias. Departamento de Ciências Sociais - ES - Brasil -
revistasimbiotica@gmail.com

INTERACCIONES HUMANO - MÁQUINA

<https://www.sciencedirect.com/topics/engineering/human-machine-interaction>

Grischa Beier, André Ullrich, Silke Niehoff, Malte Reißig , Matthias Habich (2020), señalan en su artículo de revisión la ambigüedad y vaguedad del término “***interacción***”, sin duda un concepto central en el campo de las relaciones humano – máquina. En esta perspectiva, se presentan a continuación los aportes relevantes que han realizado distintos discursos y

disciplinas a la identificación de dimensiones que resultan importantes a la hora de un acercamiento interdisciplinar a la cuestión. Precisamente, de cara a los desarrollos recientes en inteligencia artificial (IA), aprendizaje automático (ML), tecnologías de neuroseguimiento y autoseguimiento o robótica social, explorar las formas que puedan adoptar las interacciones humano-máquina (IHM) en estos campos, así como sus posibles implicaciones y consecuencias para escenarios y actividades específicos, constituye un trabajo urgente, **también para nuestro futuro en general.**

<https://doi.org/10.1057/s41599-023-02060-8> OPEN 1

Si bien existe una fuerte tradición (interdisciplinaria) de exploración de las condiciones necesarias para la interacción entre humanos y computadoras en informática, ingeniería, humanidades y ciencias, durante mucho tiempo, estas disciplinas se preocuparon principalmente por cuestiones relacionadas con interfaces de usuario apropiadas, es decir, con posibles formas de transferir datos e información de manera adecuada y efectiva entre humanos (entendidos como usuarios) y computadoras (entendidas como máquinas) con el objetivo de resolver ciertos problemas. Por lo tanto, el enfoque principal fue el desarrollo de dispositivos de entrada y salida visual, háptica y lingüísticamente adecuados para el uso efectivo de las computadoras (lo que se conoce como interacción humano-computadora (HCI).

Sin embargo, desde la década de 1980, a raíz del desarrollo de las primeras interfaces cerebro-ordenador (BCI), la realidad aumentada y virtual, el aprendizaje automático o la computación ubicua, ha habido una tendencia a alejarse de este enfoque en los dispositivos útiles hacia una comprensión más sofisticada de la interacción, a menudo refiriéndose a algún tipo de diálogo o comunicación entre humanos y máquinas en un sentido amplio. De ahí en adelante, ni se entiende al ser humano únicamente como usuario, ni se centra solamente en los ordenadores. Esto a menudo se conoce como **HMI (Interacciones Humano - Máquina)** y, en última instancia, dio lugar a los esfuerzos aún en curso para simular las características y condiciones esenciales de la comunicación humana en las máquinas.

Como ya se viene planteando, no hay acuerdo sobre qué son exactamente la conciencia, la inteligencia o la corporeidad y cómo podrían simularse de manera efectiva. Cuando se trata del desarrollo de máquinas capaces de interactuar con los humanos, se ha apelado básicamente a la sociología, la filosofía, la psicología y la ciencia cognitiva, así como a los estudios de los medios de comunicación y la ciencia de la comunicación (notoriamente, no se han tenido en cuenta las interacciones en el campo de la biología). Uno de los problemas que salió a relucir en este encuentro de discursos y disciplinas es que aunque existe una idea coloquial de lo que significa "interacción" y, por lo tanto, a qué se refiere el término, el uso de los conceptos, especialmente en los debates científicos, sigue siendo vago y ambiguo (Janlert y Stolterman, 2017: p. 105; cf. Bucy, 2004).

Los autores señalan que, "interacción" se utiliza a menudo para denotar "*acción o influencia mutua o recíproca*" (Diccionario Merriam-Webster, 2022). De acuerdo con esta comprensión, la "interacción" se refiere a: **algo** (por ejemplo, ciertas acciones) que tiene

lugar **entre dos o más entidades** (en la mayoría de los casos: humanos) **con miras a algún propósito u objetivo** (dentro de un cierto **contexto**). Y, también a que cualquiera de estas entidades asume un papel activo en el proceso de interactuar.

Ahora bien, dada esta caracterización, la razón por la que el término "interacción" es vago y ambiguo se vuelve obvia: **¿qué ocurre exactamente en las interacciones?; ¿en qué contextos?; ¿entre qué tipos de entidades?; ¿con miras a qué propósitos o metas?** Cualquiera de estas dimensiones y todas ellas en conjunto, pueden abordarse de muchas maneras diferentes y sirven para precisar de qué se está hablando en cada caso, y también para visualizar las evoluciones y transformaciones que van ocurriendo.

Así, por ejemplo, la disposición de los ordenadores para interactuar con los usuarios ("las entidades intervinientes") tradicionalmente se entendió como un intercambio directo y explícito de datos e información (qué ocurre en las interacciones) y perseguía objetivos puramente epistémicos. Es decir, las computadoras se entendían como dispositivos cognitivos "que amplían o complementan el funcionamiento cognitivo humano mediante la realización de tareas de procesamiento de información" (Brey, [2005](#): p. 384; (Dix et al, 2003; Norman, 2013)

Sin embargo, más recientemente, las interacciones humano máquina (HMI) se introdujeron cada vez más en el campo de lo ontológico. Es decir, *las computadoras [y las máquinas] simulan entornos y herramientas para interactuar con estos entornos* (Brey, [2005](#)). Por lo tanto, el análisis de estas nuevas formas de HMI implica investigar el significado de [...] acciones, [...] metas y [...] intenciones. Ya no se trata tan solo de "datos". (Müller, [2011](#)).

En este nuevo momento evolutivo, los entornos interactivos son receptivos, activos, sensibles y están en constante diálogo con las personas del entorno. En cierto sentido, los entornos en sí mismos se están volviendo más agenciales y orientados a objetivos. Debido a que la interactividad se entiende aquí como algo que requiere algún tipo de agencia, *la interactividad no se trata solo de ser reactivo y receptivo, sino también de empujar la realidad en una cierta dirección*. (Janlert y Stolterman, [2017](#)).

La disposición a estos nuevos procesos de interacción, así como los procesos (en curso), según Janlert y Stolterman, dependen de varios parámetros, como la agencia, el ritmo o el tiempo, la independencia, la receptividad, la previsibilidad y la aplicación de las reglas, que ahora entran en el foco de la informática y la informática. Los usuarios y las computadoras se comenzaron a entender cada vez más como sujetos simétricos de interacción, interactuando en los modos de comunicación y diálogo con el propósito de desarrollar y resolver ciertos problemas, así como estimular y cambiar entornos dentro de una amplia gama de contextos diversos.

En la medida en que "los sujetos" de interacción son entendidos en términos de tomadores de decisiones racionales, que interactúan en el *modo* de adoptar estrategias, la teoría de juegos se convierte en un punto de referencia importante.[3] Dicha teoría, como teoría matemática, modela situaciones de toma de decisiones en las que los participantes racionales interactúan entre sí (generalmente en situaciones de conflicto). Tales problemas o juegos interactivos de decisión involucran a "dos o más individuos que toman una decisión

en una situación en la que la recompensa para cada individuo depende (al menos en principio) de lo que cada individuo decida" (Webb, 2007).

Aproximaciones a las interacciones Humano – Máquina desde la sociología y la filosofía

En sociología, la interacción es uno de los conceptos clave que se utilizan para explicar una amplia variedad de fenómenos sociales en diferentes niveles. Su papel fundamental se deriva del hecho de que los conceptos de interacción pueden explicar fenómenos heterogéneos localizados en diferentes niveles sociales (individual, organizacional y social), como, por ejemplo, las relaciones interpersonales, la estructuración y reproducción de situaciones e instituciones sociales, así como la cohesión social y la transferencia transgeneracional de conocimientos. Sin embargo, durante mucho tiempo, el uso de la "interacción" se restringió a las relaciones entre (al menos dos) *humanos*.

Sin embargo, con la aparición de la Teoría del Actor-Red (ANT) de Bruno Latour (Latour, 2005) y las Ciencias de Tecnología y Sociedad (CTS o STC) el concepto se amplió también en lo que respecta a la interacción entre humanos y máquinas o más bien entre entidades humanas y no humanas. Así, Latour (2005), así como Castells (2010) cambiaron la perspectiva teórica hacia una perspectiva relacional, argumentando que *el mundo físico y social debe entenderse como un arreglo complejo de relaciones entrelazadas entre diferentes entidades, que no tienen que ser pensadas como estructuralmente idénticas para que se les atribuya agencia e interactividad*.

Siguiendo esta línea argumental, Latour y Castells reescriben la historia del ser humano y de la tecnología señalando que las relaciones entre el ser humano y los sistemas tecnológicos, entre el ser humano y el ser humano, así como entre el ser humano y su entorno natural, tienen la misma importancia para el análisis sociológico que las relaciones interpersonales. Siguiendo esta línea de pensamiento, Werner Rammert (2002), protagonista de la CTS alemana, planteó la cuestión de la agencia tecnológica siguiendo un enfoque funcionalista en lugar de uno materialista. Al centrarse en la función de la tecnología, Rammert demuestra la capacidad de la tecnología para actuar e interactuar en diferentes niveles sociales, dando forma a las estructuras situacionales, institucionales y sociales.

En cuanto a la cuestión de *cómo se produce la interacción*, se pueden distinguir diferentes modos de interacción. Mientras que Jensen (1998) sostiene que la interactividad y la interacción están ligadas a la adaptación mutua del comportamiento y la acción, Blumer (1986) enfatiza la importancia del intercambio mutuo de interpretaciones. Desde este último punto de vista, la interacción humana está mediada por el uso de símbolos y significados, por la interpretación o por la determinación del significado de las acciones de los demás.

La tradición del interaccionismo simbólico acentúa la importancia de la simbolización y distingue dos tipos de interacción al introducir la diferencia entre *interacción simbólica* y *no simbólica*. El primero ocurre si un compañero de interacción ya ha interpretado las acciones de su contraparte, es decir, la interacción tiene lugar en la toma de roles mutuos, así como en la interpretación del comportamiento. Por el contrario, esto último se ejemplifica en respuestas espontáneas y reactivas a las acciones de otra persona.

La interdependencia de la acción también es crucial para Starkey Duncan (1989), quien argumenta que un proceso es interactivo si al menos dos individuos interdependientes, donde la interdependencia se identifica como un estado de conciencia recíproca, es decir, cada compañero de interacción es a) consciente de la presencia del otro y b) asume que el otro es consciente de su presencia. Con respecto a *los modos* de HMI, Philip Hayes y Reddy (1983) señalan que la interacción tiene lugar si la entrada fragmentada puede ser analizada y combinada, si las habilidades y limitaciones, acciones y motivos pueden ser explicadas, y si un diálogo puede ser iniciado y perpetuado mediante el seguimiento del foco de atención.

Teniendo en cuenta los propósitos y contextos de la interacción (*¿por qué se produce la interacción? ¿Dónde se está llevando a cabo la interacción?*), el concepto de interacción apunta a la construcción y reproducción de estructuras sociales en diferentes niveles sociales. Por lo tanto, la interacción es un proceso multiestable, cuyo propósito depende del contexto específico en el que se lleva a cabo la interacción.

Siguiendo a Bahr y Stary (2016), la interacción debe considerarse principalmente como un proceso de intercambio de bienes materiales e inmateriales entre partes actuantes (entidades biológicas o técnicas) encarnadas en un contexto determinado. Este énfasis en el contexto, que es compartido por la Teoría del Actor – Red (ANT), por ejemplo, abre la perspectiva analítica para la importancia de la diferenciación y la imposibilidad de generalizar hipótesis sobre los posibles resultados de la interacción. La gama de resultados posibles incluye la creación de orden y significado (interacciones simbólicas), así como la construcción de estructuras organizacionales y sociales establecidas a través de esquemas conductuales y cognitivos que resultan de interacciones repetidas e influyen en eventos futuros como patrones y expectativas.

Por lo tanto, se deben identificar las características específicas de cada interacción, contexto y socio, ya que la misma interacción puede conducir a diferentes resultados dependiendo de las condiciones de su contexto (por ejemplo, una interacción entre un robot y un residente de un hogar de ancianos puede conducir a otro resultado que la interacción entre un cliente de un supermercado y el mismo robot). Estas diferencias deben tenerse en cuenta a la hora de desarrollar o evaluar casos concretos de HMI.

Resumiendo las diferentes corrientes teóricas, los conceptos sociológicos de interacción pueden diferenciarse en función de las condiciones previas que deben cumplir los socios de la interacción. Si bien la mayoría de las teorías reservan exclusivamente el concepto de interacción para los procesos de intercambio humano y sus efectos, con el auge de ANT y STS, se adoptó un enfoque diferente, que es de particular relevancia para el análisis y la evaluación de la HMI. Al ampliar la gama de posibles socios de interacción al ámbito de las máquinas (en el sentido más amplio de este concepto), ANT y STS permiten utilizar la "interacción" como un concepto heurístico para explicar las condiciones tecnológicamente habilitadas de la sociabilidad moderna. Por lo tanto, estos enfoques parecen especialmente fructíferos para comprender las estructuras sociales modernas y las relaciones sociales, mientras que los conceptos de interacción derivados de la esfera interpersonal solo deben transferirse con cautela a los contextos de HMI, ya que sus condiciones previas a menudo no son satisfechas por los sistemas técnicos.

En filosofía de la tecnología, el desarrollo del concepto de interacción se extiende desde la era preindustrial, pasando por la industrialización, hasta nuestros días: en la era preindustrial, la atención se centraba en los seres humanos en el "espejo de sus máquinas" (Meyer-Drawe, 1996) y, por lo tanto, en su *comparación*. Durante la industrialización, este enfoque se desplazó hacia las consecuencias sociales de una creciente integración de las máquinas en la vida cotidiana: a partir de entonces, ya no se trataba principalmente de la comparación humano-máquina, sino de la interacción humano-máquina (Gehlen, 2007; Müller y Liggieri, 2019). Con el rápido desarrollo de nuevos tipos de tecnologías, los filósofos de la tecnología se hacían cada vez más preguntas sobre las interacciones concretas entre humanos y máquinas. Este enfoque se refleja en *las consideraciones situadas y contextuales de los entornos individuales de la HMI, que toman los modos y contextos como aspectos centrales de la interacción (¿cómo tiene lugar la interacción? ¿Dónde se lleva a cabo la interacción?)*. Esto se hace particularmente claro en los enfoques recientes en el campo de la post-fenomenología, así como en la tecnociencia. Tanto Peter-Paul Verbeek como Karen Barad, como pensadores contemporáneos clave de los dos campos, abordan de manera crítica y productiva la noción de interacción centrándose en el aspecto del "entre" en HMI.

Por el contrario, Shaun Gallagher, un conocido pensador del enactivismo, introduce su *Teoría de la Interacción (TI)* en la que arroja luz sobre el contexto mismo de las interacciones concretas a partir de la tesis de que toda forma de interacción sólo puede ser determinada por su situación. Por último, hay que mencionar enfoques como la Filosofía de la Información de Luciano Floridi, de orientación analítica, que se centra en la HMI desde una perspectiva ética y epistemológica.

Verbeek, que es uno de los principales pensadores de las filas de los post-fenomenólogos, está particularmente preocupado por el fenómeno de la "mediación de la tecnología humana" (Verbeek, 2005). Por lo tanto, no se centra directamente en el concepto de interacción, sino más bien en un examen crítico de situaciones concretas en las que se revelan ciertas relaciones entre los seres humanos y las tecnologías. Por lo tanto, para Verbeek, la interacción es sólo una de las varias relaciones posibles entre los seres humanos y la tecnología. Con esto, llama la atención sobre el hecho de que en entornos prácticos concretos, es a través de la interacción que las entidades involucradas aparecen por primera vez como lo que son: los humanos y las máquinas "*no son entidades pre-dadas, sino más bien [...] se moldean mutuamente en las relaciones que se producen entre ellos*". (Verbeek, 2015). A este respecto, los análisis postfenomenológicos de la HMI se refieren, en primer lugar, al *modo* de interacción y, en segundo lugar, al contexto concreto y práctico de la HMI.

En el campo de la tecnociencia, es sobre todo Karen Barad (2007) quien se ocupa de una discusión crítica de la relación entre los seres humanos y las tecnologías (o más en general: los objetos). Su punto de partida es una crítica a la posición del sujeto dentro de las tradiciones filosóficas. Para apoyar esto, desarrolla el concepto de intra-acción, un término que se utiliza para reemplazar el concepto de "interacción", que presupone cuerpos preestablecidos que participan en acción entre sí. En contraste, la intra-acción entiende la agencia no como una propiedad inherente de un individuo o ser humano a ser ejercida, sino como un dinamismo de fuerzas en el que todas las entidades designadas están

constantemente intercambiando y difractando, influyendo y trabajando inseparablemente. De este modo, "[...] reconoce la imposibilidad de una separación absoluta u objetividad entendida clásicamente, en la que un aparato (una tecnología o medio utilizado para medir una propiedad) o una persona que usa un aparato no se consideran parte del proceso que permite 'resultados' o mediciones específicamente ubicados" (Stark, 2016). *Más bien, "[...]' individuos' solo existirían dentro de los fenómenos (relaciones particulares materializadas/materializadoras) en su continua reconfiguración iterativamente intraactiva"* (Barad y Kleinman, 2012).

En este sentido, Barad, al igual que Verbeek, por un lado, se centra en la "intermediación" ("in-betweenness") de los humanos y las máquinas, mientras que, por otro, se interesa por las formas concretas en que los agentes aparecen en las interacciones.

Aunque Shaun Gallagher no se centra explícitamente en la HMI, define la interacción de tal manera que puede utilizarse para el análisis de la HMI. El concepto de interacción de Gallagher se sitúa en *una perspectiva enactivista*, implementada en lo que Gallagher, a diferencia de la Teoría-Teoría y la Teoría de la Simulación, llama TI. Para él, "la informática enfatiza la importancia del contexto y las circunstancias, y el papel de las prácticas comunicativas y narrativas" (Gallagher, 2020). De ahí que Gallagher —en analogía con Verbeek y Barad— señale el modo y el contexto específicos de la interacción, en la medida en que el modo se expresa en el enfoque en las prácticas comunicativas y narrativas y el contexto en el enfoque en la situacionalidad de las interacciones: la interacción se presenta como "un acoplamiento corregulado mutuamente comprometido entre al menos dos agentes autónomos", distinguiendo entre "(a) la corregulación y el acoplamiento se afectan mutuamente y constituyen una organización autosuficiente en el dominio de la dinámica relacional", y "(b) la autonomía de los agentes involucrados no se destruye, aunque su alcance puede aumentarse o reducirse". (Gallagher, 2020).

En contraste con esto, la explicación analítica de Luciano Floridi sobre la interacción es, al menos, doble: primero, en su *Ética de la información*, afirma que "la interactividad significa que el agente y su entorno (pueden) actuar el uno sobre el otro" (Floridi, 2013). Es decir, los sistemas autónomos, adaptables y situados, interactúan con su entorno para transformarse o producir ciertos efectos entre sí a través del intercambio de información. De ahí que, según Floridi, no solo puedan interactuar los agentes humanos, sino también los agentes artificiales (en la medida en que sean autónomos y adaptables), por lo que ambos pueden actuar como agentes morales ("si y solo si son capaces de una acción moralmente calificable" (Floridi, 2013) Esto, por supuesto, tiene importantes implicaciones éticas al menos para ciertos contextos de interacción.

En resumen, en los campos de la (post)fenomenología, la tecnociencia y el enactivismo, está **emergiendo actualmente una conciencia de los límites y posibilidades de la noción de interacción**. En estos campos, los pensadores recurren al concepto de interacción de una manera fructífera y productiva, incorporándolo en contextos prácticos y situados para abordar modos y contextos particulares de interacción. En este contexto, los rendimientos centrales son, por ejemplo, la reflexión crítica sobre la relación entre los humanos y la tecnología o las máquinas, que pone de relieve una dependencia mutua entre

humanos y no humanos o entre la tecnología y hace hincapié en el contexto concreto en el que la "interacción" aparece como un concepto significativo.

La "interacción" en psicología y Ciencia Cognitiva

En psicología, el término "interacción" se utiliza y estudia en diferentes subdisciplinas (por ejemplo, psicología social, psicología diferencial, psicología del desarrollo, ergonomía, psicología industrial u organizacional), lo que da lugar a una amplia gama de diferentes comprensiones. En general, en lo que respecta a las dimensiones de *¿quién interactúa?* y *¿Dónde tiene lugar la interacción?*, las teorías psicológicas se refieren a dos (o más) variables, estados psicofisiológicos o interpersonales, constructos, sistemas, condiciones ambientales, personas o comportamientos que interactúan en contextos sociales específicos (Dix et al., 2003; Bolis y Schilbach, 2020). Sin embargo, las teorías psicológicas se centran más en *los posibles modos y propósitos* de la interacción (*¿cómo tiene lugar la interacción? ¿Por qué se produce la interacción?*), así como sobre las posibles *consecuencias psicológicas* de la interacción.

La pregunta general de *cómo* se produce la interacción se responde de diferentes maneras, por ejemplo, en el sentido de estar estrechamente relacionados entre sí, de intercambiar información, de interdependencia, de dirección o control recíproco, de una influencia conjunta sobre algo tercero, o de transformaciones colectivas del mundo (Hasson y Frith, 2016; Bolis y Schilbach, 2020). En general, la interacción se entiende como un proceso dinámico que se relaciona con una adaptación mutua continua, un acoplamiento dinámico de las partes que interactúan y un desarrollo relacionado de comportamiento complementario (Hasson y Frith, 2016).

Existe una amplia variedad de propuestas en la literatura psicológica con respecto a modos concretos de interacción (por ejemplo, por gesto, expresión facial, comunicación de contenido, cognición, emoción, intimidad, etc.). En cuanto a la HMI, *los aspectos cognitivos* o estados mentales (pensamiento, aprendizaje, memoria, atención, percepción, planificación, toma de decisiones, etc.) juegan un papel especialmente importante en el contexto de la psicología y la ciencia cognitiva (Sharp, 2019; Cross y Ramsey, 2021): las teorías de la cognición (por ejemplo, como ya vimos anteriormente, modelos mentales, procesamiento de la información, cognición distribuida, interacción encarnada) se consideran muy importantes para el estudio, desarrollo y diseño de máquinas interactivas (Sharp, 2019). Lo mismo ocurre con *las emociones*, que son otro componente importante de la comprensión psicológica de la interacción y de la investigación en psicología relacionada con los posibles modos de interacción: por un lado, las respuestas emocionales de los usuarios a las máquinas o a su diseño influyen significativamente en las HMI concretas. Por otro lado, la interacción con las máquinas puede utilizarse para influir en las emociones de las personas de diversas maneras (véase, por ejemplo, las tecnologías persuasivas) (Sharp, 2019).

Los modos de *interacción social*, por ejemplo, en la colaboración, la comunicación y la coordinación, también son muy importantes con vistas a la HMI. En psicología, la interacción social se entiende generalmente como un proceso bidireccional. Sin embargo, las condiciones previas para tales procesos no solo complican el desarrollo de la HMI, sino

también la comprensión psicológica de la interacción humana con las máquinas. Los enfoques más prominentes para describir las condiciones previas de la interacción social (por ejemplo, terreno *común*, *toma de perspectiva* y *Teoría de la Mente* (ToM)) asumen que los seres humanos están equipados con conocimiento directo e *implícito* de otros seres humanos (en lo que respecta, por ejemplo, a ciertas capacidades de otros humanos, sus necesidades fisiológicas, etc.).

Una dificultad clave para la HMI en comparación con las interacciones entre humanos resulta del hecho de que al menos algunos aspectos de la interacción social entre humanos nunca se hacen explícitos (Krämer et al., [2011](#)). Además, incluso si todos los aspectos de la interacción humano-humano pudieran hacerse explícitos, la implementación de las reglas o conocimientos subyacentes no sería suficiente para establecer una mentalización exitosa en las máquinas y, por lo tanto, para permitir el desarrollo de formas de HMI que sean similares a la interacción humano-humano (Frith y Frith, [2003](#); Krämer et al., [2011](#)).

Otro modelo de interacción de despliegue difícil de aplicar a las máquinas se expresa en *la teoría de la penetración social* (Krämer et al., [2011](#); Fox y Gambino, [2021](#)). De acuerdo con este relato, es necesaria una auto-revelación recíproca escalonada para desarrollar relaciones. Sin embargo, cuando las máquinas comparten información, dicha información no se basa en valores personales, experiencia o autoimagen (Fox y Gambino, [2021](#)). Por lo tanto, desde un punto de vista psicológico, hay muchos indicios de que la comprensión de la interacción social, es decir, el *cómo* de la interacción entre humanos, no es completamente transferible a la comprensión de la interacción entre humanos y máquinas.

Por lo tanto, para la conceptualización psicológica de la interacción entre humanos y máquinas y para la cuestión de *cómo* tiene lugar la interacción con las máquinas, puede ser más importante comprender, reconocer e implementar diferentes *tipos de interacciones*. Por ejemplo, los usuarios pueden dar instrucciones a un sistema (escribir comandos, gesticular, etc.) (*tipo de interacción instructiva*), entrar en diálogos con sistemas (*tipo conversatorio*), moverse a través de espacios (virtuales) (*tipo exploración*) o responder a interacciones iniciadas por el sistema (*tipo de respuesta*) (Sharp, [2019](#))

Un propósito central de la interacción humano-humano (*¿por qué se produce la interacción?*) es formar, mantener y dar forma a las relaciones. En psicología, este aspecto también se ha abordado en relación con la HMI, tanto como objetivo de diseño como con vistas a los efectos psicológicos de las relaciones entre humanos y máquinas. La motivación para formar relaciones generalmente se encuentra en la necesidad humana *de pertenecer* (Krämer et al., [2011](#); Fox y Gambino, [2021](#)): los seres humanos como seres sociales buscan la compañía de los demás, interactúan con los demás y forman diversas formas de relaciones interaccionales (familiares, relaciones de pareja, amistad, etc.). En este contexto, las interacciones con las máquinas, en principio, también pueden conducir a relaciones entre humanos y máquinas. Más interesante desde una perspectiva psicológica, sin embargo, es la cuestión de bajo qué condiciones exactas la HMI puede conducir a la formación de relaciones. Las condiciones que se discuten en este contexto son, entre otras, un cierto grado de atractivo, similitud o simpatía recíproca (Krämer et al., [2011](#); Fox y Gambino, [2021](#)).

Otros aspectos relativos a los posibles propósitos de la interacción son recogidos por *la teoría del intercambio social*, *la teoría de la equidad* o *la teoría de la inversión* (Krämer et al., [2011](#); Fox y Gambino, [2021](#)). Una idea básica del intercambio social es que las personas necesitan intercambiar diferentes bienes (tangibles o abstractos y con diversas funciones) y cooperar para sobrevivir. Las consideraciones de costo-beneficio juegan un papel crucial aquí. En consecuencia, los seres humanos preferirían relaciones que ofrezcan más ventajas que desventajas o costos (en comparación con otras relaciones), relaciones en las que hayan invertido más en comparación con otros, o relaciones en las que los socios sean iguales en el proceso de intercambio. Sin embargo, cuanto más larga y estrecha es una relación, menos parece ser la igualdad una prioridad. Estas consideraciones también se aplican a las interacciones humanas con las máquinas, analizando cómo las personas utilizan las compensaciones de costo-beneficio para entrar en interacciones y relaciones satisfactorias con los robots, por ejemplo.

En resumen, según la comprensión psicológica, la interacción no es un fenómeno exclusivamente humano. Así, no solo los humanos interactúan entre sí, sino también, por ejemplo, con ciertas características humanas (comportamiento, gestos, cognición, emociones, etc.), sistemas, condiciones ambientales o máquinas. Además, los tipos de interacción juegan un papel importante en la descripción y comprensión de los diferentes modos de interacción. En este contexto, las características cognitivas, emocionales y sociales de los seres humanos son decisivas para la descripción psicológica y el análisis de los modos HMI. Sin embargo, si la interacción social entre humanos y máquinas se entiende por referencia a las teorías de las interacciones entre humanos, los límites de la transferibilidad se hacen evidentes rápidamente.

"Interacción" en los estudios de medios y las Ciencias de la Comunicación

Con la entrada del ordenador personal en los procesos cotidianos, los estudios de medios y comunicación experimentaron un repunte radical. En la década de 1980, por ejemplo, se estableció el campo de la Teoría de los *Medios Alemana* que, bajo los auspicios de Friedrich Kittler, identificó las operaciones de transmisión, almacenamiento y procesamiento como funciones básicas de los medios (Kittler, [1993](#)). Mientras que los estudios de medios alemanes se desarrollaron a partir de las humanidades (más precisamente, los estudios literarios) y tienen un enfoque histórico y teórico, los estudios de comunicación tienen una orientación más empírica, es decir, estudian el uso y la aplicación de las tecnologías de los medios y realizan investigaciones sobre los medios. Especialmente en este último enfoque, los conceptos de interactividad (que se refieren tanto a los procesos de interacción como a sus condiciones necesarias) juegan un papel central.

Desde el relato de Sheizaf Rafael ([1988](#)) sobre la comunicación reactiva, por ejemplo, la interacción se ha entendido como una forma de pensar sobre la comunicación, donde originalmente se asumía como un atributo de la conversación cara a cara (Isótaló, [1998](#); Rafaeli y Sudweeks, [1998](#)) que luego se extendió a los entornos de comunicación mediada y, como tal, debe distinguirse de la interacción social (Bucy, [2004](#)). En este último sentido, la interacción en los estudios de medios y la ciencia de la comunicación "se utiliza como un concepto amplio que abarca los procesos que tienen lugar entre los receptores, por un lado, y un mensaje mediático, por el otro" (Jensen, [1998](#): p. 188).

En este contexto, Bucy (2004) distingue dos tipos principales de conceptos de interacción: en primer lugar, los conceptos que se centran en la interacción humana con ciertos contenidos (por ejemplo, en línea) y abordan el control que los usuarios ejercen sobre su selección y presentación (cf., por ejemplo, Steuer, 1995; McMillan, 2002; Stromer-Galley, 2004); y en segundo lugar, los procesos de interacción que implican conversaciones de persona a persona mediadas por la tecnología (cf., por ejemplo, Massey y Levy, 1999).

Además, algunos enfoques tratan de combinar estos dos tipos centrándose en el aspecto del control en los procesos de interacción mediada de persona a persona (cf., por ejemplo, Williams et al., 1988; Neuman, 1991). Estos *enfoques centrados en el mensaje tienen en común* su fuerte enfoque en los usuarios o, en general, su enfoque en los remitentes y destinatarios humanos. En los enfoques centrados en el mensaje, los usuarios y los medios de comunicación se consideran sujetos asimétricos de interacción con un enfoque en los usuarios (refiriéndose a la cuestión de *quién interactúa*), que seleccionan, presentan y controlan ciertos contenidos (refiriéndose a la cuestión de *cómo se produce la interacción*) con el propósito de comunicar (refiriéndose a la pregunta de *por qué tiene lugar la interacción*) en contextos mediatizados (refiriéndose a la cuestión de *dónde tiene lugar la interacción*).

Por el contrario, *los enfoques estructurales* consideran el atributo tecnológico o característica de los medios, que permite a los usuarios hablar con otros usuarios, interactuar con los medios de comunicación o manipularlos, o influir en su contenido. Aquí, la interactividad se ubica como una propiedad de la tecnología o los medios.

En este sentido, Jensen, por ejemplo, afirma que la interactividad sería "una medida de la capacidad potencial de un medio para permitir que el usuario ejerza una influencia sobre el contenido y/o la forma de la comunicación mediada" (Jensen, 2008). Con este fin, Jensen distingue tres formas principales de definir la interacción en los estudios de medios y en la ciencia de la comunicación: en primer lugar, los enfoques que definen la interactividad a través de ejemplos prototípicos (refiriéndose, por ejemplo, al teléfono, los sistemas de audioconferencia o el correo electrónico). En segundo lugar, enfoques que definen la interactividad a través de ciertos criterios que se consideran necesarios para un diálogo recíproco entre usuarios y sistemas. Y tercero, la comprensión de la interacción como un continuo que puede estar presente en diversos grados

En los enfoques estructurales los usuarios y los medios de comunicación se entienden como sujetos asimétricos de interacción con un enfoque en los medios y la tecnología, lo que permite a los usuarios influir en ciertos contenidos con el propósito de comunicarse en contextos mediados.

Aparte de los enfoques estructurales, *los enfoques perceptivos* (cf., por ejemplo, McMillan, 2002) consideran las percepciones de los usuarios como unidad de medida: el grado de interactividad, que ahora se supone que tiene efectos variables, se refleja en el grado en que los usuarios experimentan subjetivamente la interactividad. En los enfoques perceptuales, los usuarios y los medios se consideran sujetos asimétricos de interacción

con un enfoque en los usuarios y su experiencia de interacción con los medios y la tecnología con el propósito de comunicarse en contextos mediados.

Kiousis, por ejemplo, combina un enfoque estructural y perceptual al postular que "[l]a interactividad puede definirse como el grado en que una tecnología de la comunicación puede crear un entorno mediado en el que los participantes pueden comunicarse (uno a uno, uno a muchos y muchos a muchos), tanto sincrónica como asincrónicamente, y participar en intercambios recíprocos de mensajes [...]. En cuanto a los usuarios humanos, también se refiere a su capacidad para *percibir* la experiencia como una simulación de la comunicación interpersonal y aumentar su conciencia de la telepresencia" (Kiousis, 2002: p. 372).

Sin embargo, en la ciencia de la comunicación tradicional, los enfoques estructurales se discutieron de manera bastante crítica. Ha y James (1998: p. 461), por ejemplo, afirman que "[l]a interactividad debe definirse en términos de la medida en que el comunicador y la audiencia responden a, o están dispuestos a facilitar, las necesidades de comunicación del otro" y, por lo tanto, afirman que los enfoques centrados en el mensaje son las únicas explicaciones plausibles de la interacción. Schumann et al. justifican esta afirmación postulando que "íntimamente es la elección del consumidor interactuar, por lo que la interactividad es una característica del consumidor y no una característica del medio. El medio simplemente sirve para facilitar la interacción" (Schumann et al., 2001: p. 45).

De nuevo un breve salto hacia los sistemas vivos..

Dando de nuevo el salto hacia los sistemas vivos en general y cómo éstos se comunican entre sí, logramos salir del antropocentrismo por la otra puerta. ..."Los sistemas vivos no se comunican simple y llanamente con base en signos y símbolos como los seres humanos, y como las explicaciones que abierta o tácitamente se fundan en la comunicación entre los seres humanos podrían llevar a pensar. Por el contrario, la forma de base de la *comunicación en la naturaleza es del orden geométrico, esto es, literalmente, a partir de formas, estructuras, mapas, configuraciones*, y no en términos de objetos y relaciones entre objetos. "... (Mc.Cabe, 2014 en Maldonado, 2016)

Aquí la ciencia que comienza a darnos otras pistas en torno a las interacciones en los sistemas vivos es la química. (Por ejemplo, conceptos tales como quimiotaxis positiva). Precisamente, la hypercomputación biológica consiste en comunicación geométrica y la síntesis basada en formas, patrones y estructuras constituye el modo de interactividad; algo que resulta familiar para la actividad artística.

Nos encontramos, pues, en medio de un asunto evolutivo de enorme importancia: La ciencia y la tecnología han dado lugar a formas de vida y de inteligencia que saben hacer muchas cosas mejor que los seres humanos. Se trata de vida y de inteligencia artificial que aprende, aprende muy rápido y aprende por sí mismas. ¿Podrán los seres humanos coexistir y convivir con la Inteligencia Artificial y con la Vida Artificial? ¿Qué tendremos que aprender nosotros para darle forma a este convivio? ¿Cómo lo imaginamos?

Como plantea Maldonado (2023), el convivio siempre ha sido un problema crítico para los humanos y específicamente, para la civilización occidental. El convivio con otros humanos y el convivio con la naturaleza, esto es, con otros sistemas vivos e incluso con nuestro propio cuerpo, han mostrado serios límites. Con toda seguridad, las tres leyes de la robótica propuestos por Asimov han quedado obsoletas...

Algoritmos: Instragram, Facebook, Tick Tock son algoritmos. Algunos nos atrapan al ir presentándonos esas cosas que más nos interesan. Hoy en día son tan importantes que ya no pueden considerarse como una parte o un simple método lógico para pensar en informática, sino que constituyen *una tecnología; son el producto*. Existen empresas que surgen porque han creado un algoritmo que se diferencia en su operatividad de otros.

No se sabe cómo funcionan los algoritmos ni qué son en realidad, dando pie a un debate ontológico. Es como si en cierto sentido cobraran "vida propia".

Si bien se basan en un input muy acotado y una salida acotada, el modelo no se construye solamente con datos externos; opera mediante "pesos diferenciales" en una red, a la manera de un algoritmo empaquetado.

La comprensión de los algoritmos sigue siendo sorpresiva...

[2] Teoría computacional de la mente: la revolución cognitiva y el nacimiento de las Ciencias Cognitivas fueron influenciados fuertemente por el desarrollo de las computadoras. En 1943, el neurólogo McCulloch y el neurofisiólogo Pitts sugirieron que la actividad del sistema nervioso central y las operaciones en las redes neuronales podrían ser explicadas mediante computaciones neuronales que siguen los principios de la lógica, con base en proposiciones que llevan a un todo (activación) o nada (ausencia de activación). Ambos investigadores son considerados unos de los primeros teóricos que concibieron los fundamentos de la computación neuronal. Su trabajo inspiró el desarrollo posterior de las redes neuronales artificiales, generando un gran interés en los computólogos para desarrollar algoritmos e implementaciones que usaran el cerebro y su funcionamiento como inspiración. A su idea se le puede atribuir el *origen de la teoría computacional*

de la mente, la cual sugiere que la mente puede ser entendida como un sistema de procesamiento de información gobernado por la manipulación formal de símbolos.

<https://www.cienciascognitivascorporizadas.com/historia/met%C3%A1fora-de-la-mente>

[3] Teoría de Juegos: desde la demostración del teorema *mínimo-máximo* de John von Neumann (1928), la teoría formal moderna de juegos sirve para múltiples campos de aplicación, por ejemplo, en economía (Laffont, 1997), sociología (Swedberg, 2001), filosofía (de Bruin, 2005), biología (Tomlinson, 1997) y, en particular, en HMI (Li et al., 2019).



ORGANIZA:



COSMOGONIANDO
creando lugares dentro de lugares