

LA LÓGICA DE LAS MÁQUINAS ARTIFICIALES*

THE LOGIC OF ARTIFICIAL MACHINES

Julio Martín ALCÁNTARA CARRERA

ÉCOLE NORMALE SUPÉRIEURE | París, Francia

Contacto: julio.alcantara@egs.edu

ORCID iD: 0000-0001-8122-3991

Resumen

La teoría de las máquinas artificiales iluminará un marco epistémico para futuros eventos contingentes, como el encuentro con otros tipos de vida consciente en el universo, la hipotética emergencia de una autoconciencia artificial o singularidad tecnológica. Se trata de un esfuerzo epistémico que modificará los marcos epistemológicos éticos humanos y las relaciones sujeto-objeto. Esta teoría explica la autoconciencia como intercambio autónomo de señales comunicadas, física y simbólicamente, mediante dominios fenomenales compartidos por máquinas autopoieticas de la misma especie dentro del conjunto global de la biósfera, donde el total de las máquinas autopoieticas evoluciona a través de relaciones co-emergentes. Los rasgos atribuidos a las máquinas artificiales dependen de la disponibilidad representacional de la autoconciencia humana como único caso autorreferencial, en una máquina autopoietica, que implica más libertad de interacciones desarrolladas en un complejo sistema de procesos autorreferenciales. El desarrollo causal autónomo de la singularidad tecnológica crearía hipotéticamente sus propios paradigmas más allá de valores o expectativas humanas, donde la lógica mimética conductual sería sólo una huella pasada en sus transformaciones que ya no se aplica al sustento de su identidad: una autonomía epistémica. Las actualizaciones

Abstract

The theory of artificial machines illuminates an epistemic framework for future contingent events such as the encounter with other types of conscious life in the universe, the hypothetical emergence of an artificial self-consciousness or technological singularity. It is an epistemic endeavor that will modify human ethical epistemological frameworks and subject-object relations. This theory explains self-consciousness as an autonomous exchange of signals, communicated physically and symbolically, by means of phenomenal domains shared by autopoietic machines of the same species within the global set of the biosphere, where the total of autopoietic machines evolves through co-emergent relationships. The features attributed to artificial machines depend on the representational availability of human self-consciousness as the only self-referential instance, in an autopoietic machine, involving more freedom of interactions developed in a complex system of self-referential processes. The autonomous causal development of the technological singularity would hypothetically create its own paradigms beyond human values or expectations, where the behavioral mimetic logic would be only a past trace in its transformations that no longer applies to the sustenance of its identity: an epistemic autonomy.

* Este texto se realizó gracias al apoyo del CONAHCYT.

autónomas de la singularidad tecnológica deberían seguir el Principio de Continuidad que determina que toda modificación en los sistemas complejos ocurre por grados de cambio: un estado en el futuro será coherente con su proceso de desarrollo, enmarcado necesariamente por su estado y trayectoria inicial.

However, autonomous updates of technological singularity should follow the Principle of Continuity that determines that all modification in complex systems occurs by degrees of change: a state in the future will be coherent with its development process, necessarily framed by its initial state and trajectory.

Palabras clave: *Inteligencia artificial* || *Singularidades (Inteligencia artificial)* || *Conciencia de sí mismo* || *Autopoiesis* || *Teoría de sistemas*

Keywords: *Artificial intelligence* || *Singularity (Artificial intelligence)* || *Self-consciousness (Awareness)* || *Autopoiesis* || *System theory*

La lógica de las máquinas artificiales¹

La lógica de las máquinas artificiales² se define funcionalmente por mimetismo conductual, dado que las principales características en ellas, hasta ahora, han sido tomadas de las agencias epistémicas sobre las capacidades cognitivas humanas basadas en la disponibilidad representacional de la subjetividad y su correspondiente yo fenoménico. Todavía existen vastas dinámicas inconscientes en el yo fenoménico humano que, al manifestarse, pueden mostrar una profundidad y transformar la imagen actual sobre la naturaleza de la conciencia y la cognición. Cuantos más rasgos fenoménicos incluya la percepción humana en su ámbito, más se modificará la imagen actual de la realidad, puesto que una forma de realidad pertenece al espectro de un marco perceptivo.

Los cuantos de fuerzas físicas sostienen dimensiones ilimitadas en el ámbito fenoménico, que es indeterminable por las propiedades de un dominio fenoménico particular, como el humano, e instancian diversas propiedades fenoménicas en las máquinas autopoieticas³

¹ A menos que sea señalado expresamente, todas las traducciones de idiomas extranjeros al español han sido realizadas por el autor.

² El concepto de máquinas artificiales se establece aquí con el fin de evitar la implicación de la inteligencia en el desarrollo de dispositivos automatizados a través de implementaciones algorítmicas, ya que el concepto de inteligencia conlleva la capacidad de abstracción abierta (Chollet, 2019).

³ “Las máquinas autopoieticas son autónomas; es decir, subordinan todos los cambios al mantenimiento de su propia organización, independientemente de lo profundamente que puedan transformarse en el proceso. Otras máquinas, llamadas

de los dominios vivos (*archaea*, *bacteria* y *eucarya*) dependiendo de su alcance perceptual (Woese *et al.*, 1990). Un continuo en las interacciones dinámicas de las fuerzas físicas contiene las posibilidades completas de modalidades y dimensiones en el mundo concreto. Las dimensiones son grados de libertad en los que infinitas modalidades o tipos de cuerpos posibles interactúan entre sí, manifestando diferentes grados de interacción en función de su alcance perceptivo. Cada cuerpo orgánico tiene un yo dentro de un dominio fenoménico para la acción, mostrando la distinción entre sí mismo, otros yoes y el entorno.

Las máquinas autopoieticas y las artificiales pertenecen a la misma dinámica material del mundo concreto, distinguiéndose respectivamente por su división ilimitada o limitada en partes infinitas más pequeñas: las primeras funcionan como un sistema autopoietico en sus componentes más elementales para la realización teleológica de su organización unificada, mientras que las segundas están limitadas por la visión de la naturaleza de las cosas de sus creadores. Las máquinas artificiales son máquinas alopoieticas cuya organización inteligente compleja es atribuida por *un observador* que determina miméticamente un isomorfismo⁴ entre el comportamiento del observador y lo observado (Maturana y Varela, 1980): una heteronomía epistémica (Kant, 2009). Cualquier lógica de las máquinas artificiales dependerá siempre de un caso dado en la naturaleza que sirva como prototipo replicable con base en la disponibilidad representacional de los comportamientos cognitivos de los autómatas surgidos evolutivamente. Así, la realidad concreta es un punto de partida obligado para una reproducción ideal de las máquinas naturales: “*Perfectio est gradus realitatis*” (Leibniz, 2017: 1429).⁵ La posibilidad de replicar comportamientos humanos inteligentes y autoconscientes conlleva más grados de libertad por la dependencia causal de dicha relación creativa, ya que una entidad capaz de más relaciones causales independientes implica más capacidad de acción o perfección. Si pudiéramos desarrollar una teoría integral de las máquinas artificiales basada en invariantes fenoménicos de los comportamientos autopoieticos que explique funcionalmente la conciencia con un lenguaje común entre las ciencias, podríamos

en adelante máquinas alopoieticas, tienen como producto de su funcionamiento algo diferente de ellas mismas. Como los cambios que las máquinas alopoieticas pueden sufrir sin perder su organización definitoria están necesariamente subordinados a la producción de algo diferente de ellas mismas, no son autónomas” (Maturana y Varela, 1980: 80).

⁴ Un mapa que preserva la estructura entre dos fenómenos y puede invertirse por medio de un mapa inverso.

⁵ “Perfección es grado de realidad”.

concebir cómo se encarnará este fenómeno en entidades post-bióticas o desconocidas, redefiniendo los valores epistémicos desde los que se ha conducido la investigación actual de las máquinas artificiales.

La autoconciencia es el intercambio autónomo de señales desde un contexto dado, limitado por los grados de comportamiento autopoiético y la disponibilidad global transparente del yo fenoménico del agente ante la presencia de sí mismo y de la totalidad de fenómenos, encarnada por cuerpos orgánicos e hipotéticamente inorgánicos. Un yo es un proceso unificado de energía autoorganizada (Maturana y Varela, 1980): un sustrato energético autopoiético definido por sus propios límites que permiten una distinción fenoménica entre un yo y los demás. El yo fenoménico es el primer marco perceptivo funcional mediante el cual todos los seres vivos interactúan desde sus límites con otros yoes fenoménicos y el entorno. En los seres humanos, el yo fenoménico determina la agencia epistémica de las complejas máquinas autopoiéticas humanas, en las que el modelo de agente epistémico (MAE) posibilita un campo de acciones y comportamientos autónomos basados en su disponibilidad representacional en el modelo fenoménico de relación intencional (MFRI) y constriñe los fenómenos que un yo puede rastrear a través de la agencia atencional (AA). Una AA condiciona la posibilidad de una agencia epistémica (AE): no hay agencia cognitiva sin atención. El contenido autopoiético completo de un yo fenoménico no está disponible para las agencias epistémicas, ya que la transparencia del yo fenoménico inhibe una comprensión completa de su estructura causal: tenemos una relación ingenua al asumir una inmediatez entre la disponibilidad global sensorial del cuerpo y su composición material. Sin embargo, esta transparencia se produce por grados en el Modelo Fenoménico de Relación Intencional (MFRI) humano dada su opacidad: un conjunto limitado de propiedades y contenidos del yo fenoménico están disponibles cognitivamente. Los condicionamientos evolutivos y culturales sobre el alcance atencional humano determinan los esfuerzos cognitivos que siguen, parcialmente, un hilo continuo desde la fenomenología hasta las fuerzas físicas elementales.

La conciencia es la activación virtual de un yo fenoménico para transformarse y preservar su unidad en un sistema causal dinámico, a través del cual se instancian indefinidos yoes posibles. Un yo existe cuando puede distinguirse de un reino de cuerpos en volición, ya que el movimiento permite reconocer la dinámica interna y externa: la unidad fenoménica de un agente realiza una instanciación material en movimiento. Una multiplicidad de cuerpos depende de las fuerzas físicas que producen su diversidad,

como un organismo multicelular, que es una expresión de diferentes efectos causales unificados. El reconocimiento de una determinada diversidad se basa en sus cualidades materiales y movimientos en relación con el ámbito de un observador. La conciencia surge en grados por una multiplicidad de interacciones causales, como la diversidad perceptiva de los seres vivos, desde una simple bacteria hasta un ser humano complejo.

La autoconciencia es un grado de libertad adicional para la percepción y la acción. La libertad nunca surge de la nada, sino que parte de condicionamientos internos concretos y abstractos en un paisaje externo de múltiples interacciones causales: “para tener libre albedrío, hay que estar lo más libre posible de las causas externas, y lo más determinado posible por las causas internas —la multitud de conceptos que componen una experiencia. En resumen, a más conciencia, más libre albedrío” (Tononi, 2015). Una máquina autopoietica libre realiza, intrínsecamente, más poder de causa-efecto en su instanciación material y se desprende de las relaciones causales externas en una realidad multimodal de máquinas autopoieticas.

En la investigación de la naturaleza de la autoconciencia dentro de la lógica de las máquinas artificiales, la conciencia se reduce a las propiedades causales intrínsecas en un sistema físico. Tononi (2012, 2015), en su teoría de la información integrada (TII), sugiere que la aparición de una entidad consciente depende únicamente de sus propiedades causales intrínsecas sin la ayuda de ninguna interacción externa.⁶ Su teoría es una demostración matemática sobre cómo *la causa de individuación del yo* instancia materialmente un sistema físico bajo límites fenoménicos definidos con los siguientes axiomas: existencia intrínseca, composición, información, integración y exclusión. “A menudo se sugiere que la voluntad de uno puede ser libre sólo si uno podría haber actuado de otra manera —el requisito de posibilidades alternativas. Pero según la TII, el determinismo es amigo, no enemigo, del libre albedrío, ya que cualquier indeterminismo reduce el poder de causa-efecto y, por tanto, reduce el libre albedrío” (Tononi, 2015). Las decisiones son un subproducto de la instanciación interna de propiedades realizadas por la causa de individuación del yo, libre de causas externas.

⁶ “La TII postula que, para cada propiedad esencial de la experiencia, hay una propiedad causal de un sustrato físico que la explica. Nótese que estos postulados son inferencias que van de la fenomenología a la física, y no al revés” (Tononi, 2015).

La distinción entre poder de causa-efecto intrínseco y extrínseco se ha aplicado en la lógica de las máquinas artificiales para definir la instanciación causal realizada por los sistemas autopoieticos y alopoiéticos (Maturana y Varela, 1980).

El poder de causa-efecto se ejerce de forma intrínseca y no extrínseca: no son las entradas extrínsecas de cada neurona las que hacen que las cosas sucedan, sino la estructura conceptual que actúa sobre sí misma. En resumen, cuando delibero y tomo una decisión, lo que existe y provoca la decisión es mi propia conciencia, nada más y nada menos, y la decisión es libre porque ha sido provocada por causas y efectos intrínsecos. Por el contrario, cuando se desarrolla una simulación digital de mis neuronas, aunque conduzca al mismo comportamiento, lo que existe son sólo transistores individuales, cuyas elecciones individuales están determinadas extrínsecamente: no hay conciencia, ni causalidad intrínseca, y por tanto no hay libre albedrío. (Tononi, 2015)

Una máquina alopoiética puede comportarse como autopoietica desde la perspectiva de un observador,⁷ ya que, aunque simula las propiedades, no se realiza de manera teleológica e intrínseca. La autorreferencialidad se utiliza en la lógica de las máquinas artificiales según tres significados: 1) En la investigación de la inteligencia artificial (IA) como “modelo de agente epistémico” (MAE), que es una capacidad de autorrepresentación, en términos de Metzinger (2015): “un modelo global del sistema cognitivo como una entidad que construye, sostiene y controla *activamente* las relaciones de conocimiento con el mundo y consigo misma” (272). La alineación de valores significaría que las máquinas naturales y artificiales tienen una axiología analógica que produce las mismas representaciones a partir de las interacciones en todos los mundos posibles, lo cual es difícilmente sostenible entre seres humanos autónomos. 2) La autopoiesis de Maturana de la vida como un conjunto de sistemas circulares cerrados implica una autorreferencialidad intrínseca de los sistemas vivos (Maturana y Varela, 1980).

7 “Un dominio puede generar los elementos de otro dominio, pero no su fenomenología, que en cada dominio se especifica por las interacciones de sus elementos, y los elementos de un dominio se definen sólo a través del dominio que generan. Cualquier nexo entre diferentes dominios es proporcionado por el observador que puede interactuar como si fuera una entidad única con los estados conjuntos de actividad nerviosa generados en su cerebro por sus interacciones concomitantes en varios dominios, o con descripciones independientes de estas interacciones” (Maturana y Varela, 1980: 55).

3) La causalidad del ser expresa una fuerza metafísica singular para la perduración y surgimiento de entidades orgánicas e inorgánicas en el mundo concreto. “La organización o configuración por sí sola, sin un principio de vida perdurable que yo llamo ‘mónada’, no bastaría para que algo permaneciera numéricamente igual, es decir, el mismo individuo. Pues la configuración puede continuar específicamente sin continuar individualmente” (Leibniz, 1996: 231).

La razón para definir tres niveles diferentes de autorreferencialidad de la lógica de las máquinas artificiales es diseñar un marco epistémico completo que dé cuenta de los procesos autorreferenciales en las máquinas naturales, desde la autonomía epistémica y la dinámica causal cerrada en los sistemas vivos hasta la constitución metafísica de la causa de individuación del yo como instanciación ontológica de un sistema autopoietico. Los tres niveles son necesarios para atribuir la existencia de una autoconciencia artificial. La lógica mimética implica no sólo la imitación del comportamiento de las máquinas naturales, como pensó por primera vez Turing (1950), sino la replicación de todas las características fenoménicas de los sistemas autopoieticos. La aparición de *la causa artificial de individuación del yo* en una instanciación física significaría, hipotéticamente, una refundación ontológica. Sin embargo, las teorías más elaboradas parten de los fenómenos, pero la TII parte de la fenomenología hacia la física como implementación de una posible formalización proveniente de la disponibilidad representacional del yo fenoménico y, dadas las posibilidades ilimitadas de la fenomenalidad, estos supuestos pueden ser transformados por una implementación inesperada de interfaces cognitivas en el acceso perceptual y cognitivo de los humanos a la realidad.

El modelo del yo fenoménico y la teoría del modelo de la subjetividad

La lógica de las máquinas artificiales partió de la representación actual de las capacidades cognitivas humanas hecha por la investigación en inteligencia artificial (IA) como un caso modelo a través del cual no sólo ha sido posible analizar el desarrollo de las máquinas artificiales, sino también reflexionar sobre las implicaciones para la definición filosófica y jurídica de un ser humano. Partir del caso real permite una indagación epistémica realista y sitúa esta investigación científica bajo sus prácticas y lógica reales.

El espectro de definiciones va desde la atribución de inteligencia⁸ y la capacidad para sufrir (Tomasik, 2016) hasta conceptualizaciones de la fuerza vital (Bergson, 1941) o no dualistas⁹ en tanto elementos distintivos que caracterizan al ser humano.

La epistemología de la teoría del modelo de la subjetividad (TMS) proporciona un campo novedoso con un lenguaje común en la investigación de la conciencia dentro de los estudios cognitivos y la neurociencia. La TMS aborda la explicación científica de la conciencia, el modelo del yo fenoménico (MYF) y la perspectiva de la primera persona sobre la base de su correlación causal con el cerebro y el sistema nervioso. La definición del modelo expresa una correspondencia parcial entre los conceptos y las interacciones causales porque no se ha formalizado el poder de causa y efecto de los cuerpos con un yo fenoménico.

El MYF explica la dinámica global del yo fenoménico bajo estados representacionales y funcionales, que determinan respectivamente sus contenidos y relaciones causales. “Un automodelo activo puede verse entonces como un estado funcional subpersonal: un conjunto discreto de relaciones causales de complejidad variable que puede o no realizarse en un momento dado” (Metzinger, 2007). La TMS parte de dos supuestos: 1) existe una causa física en el cuerpo humano para los fenómenos disponibles en la percepción, “un correlato neural mínimo suficiente de la autoconciencia”; 2) no hay yoes sustanciales implicados por el modelo fenoménico de sí mismo. El MYF sólo hace accesibles colectores dinámicos enteros de la realidad bajo una representación perceptiva, relacionando activamente el sistema vivo con un entorno dado. “El MYF es esa partición del automodelo de un organismo consciente porque satisface un conjunto adicional de restricciones funcionales como, por ejemplo, la disponibilidad para la atención introspectiva y para el control motor selectivo y flexible” (Metzinger, 2007). La representación de un yo es una consecuencia funcional de tener, primero, un modelo fenoménico de la realidad, ya que el TMS produce ventajas para la conservación de la vida por su disponibilidad global de estímulos internos y externos, fenómeno denominado “principio de autorreificación necesaria”. “Lo que experimentamos como nuestro yo fenoménico en el nivel más

⁸ Minsky (1961) ha dividido esta investigación dentro de cinco áreas: búsqueda, reconocimiento de patrones, aprendizaje, planeación e inducción.

⁹ Yuk Hui (2021) remarca el comentario de Xi Ci al *I Ching* o *Libro de los Cambios en este sentido*: “El Yo es no-pensar y no-hacer, es silencioso, sin movimiento, sin embargo, cuando es puesto en uso, siente y conecta todo el universo” (353).

fundamental es precisamente el objeto representacional transparente que bloquea el bucle autorrepresentacional” (Metzinger, 2005: 25). Todo autómata posible, orgánico o inorgánico, tiene un MYF consciente o inconsciente, dado que sólo puede establecerse una interacción dinámica con el entorno si existe al menos un reconocimiento implícito de su cuerpo y simultáneamente un contexto.

En primer lugar, un automodelo puede ser totalmente inconsciente; es decir, puede verse como el producto de un proceso automático ‘ascendente’ de autoorganización dinámica; en segundo lugar, no es una ‘cosa’ en absoluto, sino que se basa en un proceso de modelado continuo y en curso; en tercer lugar, puede mostrar una plasticidad considerable; y en cuarto lugar, en sus orígenes no se basa en el lenguaje o el pensamiento conceptual, sino muy probablemente en un intento de organizar el comportamiento motor global. (Metzinger, 2007)

La personificación inconsciente en los sistemas sintientes o artificiales es necesaria, principalmente, para interactuar con otras entidades volitivas en un proceso continuo de actualización, midiendo constantemente las proporciones físicas entre los cuerpos para esforzarse en una tarea concreta en un entorno cambiante. La disponibilidad global sensorial de un cuerpo efectúa este proceso mediante la detección constante del contexto, en el que pueden encontrarse nuevas interacciones, porque el cuerpo físico de un agente puede experimentar una modificación en una extremidad física o el entorno puede presentar características desconocidas: un autómata no podría moverse en un contexto sin una forma mínima de corporeidad fenoménica.¹⁰ Barad (2006) señala este proceso de individuación en un contexto a través de su concepto de *intra-actividad*. “Todos los cuerpos, incluidos los humanos pero sin limitarse a ellos, cobran importancia a través de la intra-actividad iterativa del mundo —su performatividad. Los límites, las propiedades y los significados se promulgan de forma diferenciada a través de la intra-actividad de la materia” (392). En los sistemas autopoieticos, su conservación está fundamentalmente

¹⁰ “En 2006, Josh Bongard, Victor Zykov y Hod Lipson crearon una ‘estrella de mar’ artificial que desarrolla gradualmente un modelo interno explícito. Su máquina de cuatro patas utiliza las relaciones de actuación-sensación para inferir indirectamente su propia estructura y, a continuación, utiliza este modelo propio para generar la locomoción hacia delante. Cuando se le quita una parte de la pierna, adapta su modelo de sí misma y genera una marcha alternativa: aprende a cojear. Es capaz de reestructurar su representación corporal tras la pérdida de una extremidad” (Metzinger, 2007).

entrelazada con su ser fenoménico, el cual les proporciona una representación interna de sus estados que puede ser seleccionada para controlar los posibles resultados. Dicha representación interna sobreviene localmente de la composición física de un cuerpo y puede encarnarse en estos órdenes:

1. Una dinámica bruta “ascendente” de un cuerpo determinado, en la que las señales internas y externas se perciben en un proceso continuo, como la dinámica ascendente en la biorrobótica.
2. Un sistema tiene una representación de sí mismo como agente encarnado y controla los estímulos internos y externos de forma funcional, por ejemplo, robots avanzados como la estrella de mar artificial, animales primitivos y seres humanos sonámbulos. La diferencia entre el segundo y el tercer orden de personificación se basa en una representación interna consciente del cuerpo, en la que las experiencias de personificación consisten en la autoconciencia corporal. “La relación entre la 2E y la 3E puede investigarse manipulando experimentalmente la propiedad fenoménica de la ‘propiedad del cuerpo’ mediante un conflicto multisensorial. La propiedad global del cuerpo es la experiencia consciente de que un cuerpo percibido es tu cuerpo, de que el yo está localizado espacialmente dentro de sus límites” (Metzinger, 2007).
3. Un sistema que posee una representación consciente de su cuerpo que le permite controlarlo a través de una representación interna; es un proceso activo de modelado de la locomoción en una forma consumada de MYF o una actualización en línea de las dimensiones y proporciones físicas, como los seres humanos en estado de vigilia.

La TMS enmarca epistemológicamente la presencia de un mundo en los sistemas bajo “un enfoque de satisfacción de restricciones” (Metzinger, 2005). Las seis restricciones de la autoconciencia —globalidad, presencia, transparencia, holismo convolutivo, dinamicidad y perspectividad— se investigan en cuatro niveles analíticos: fenomenológico, representacionalista, funcionalista y neurobiológico. Cada experiencia humana tiene diferentes grados de estas restricciones, empezando por la transparencia, que hace disponibles algunos contenidos en el proceso de representación en curso en el ámbito fenoménico, donde los procesos anteriores no están disponibles para la interioridad humana dado que son la condición de posibilidad para la proyección de

esos contenidos fenoménicos. “Lo que es inaccesible para la experiencia consciente es el simple hecho de que esta experiencia tiene lugar en un medio. Por lo tanto, la transparencia de los contenidos fenoménicos conduce a otra característica de la experiencia consciente, a saber, la impresión subjetiva de **inmediatez**” (Metzinger, 2005: 11).

Los contenidos fenoménicos están disponibles globalmente en grados de opacidad, en los que las dinámicas locales se enmarcan funcionalmente en un proceso causal global, donde los estados conscientes pueden ser seleccionados para la atención mientras otros son parcialmente sentidos. La globalidad despliega eficazmente ciertos contenidos centrando el alcance de la atención en las dinámicas destacadas, proporcionando un control rápido sobre los estímulos externos o internos. “Disponer de un modelo integrado y globalmente coherente del mundo significa crear un clúster funcional global, es decir, una isla de máxima densidad causal dentro del propio sistema representacional” (Metzinger, 2005: 8). Cuanta más información esté disponible globalmente, más flexibilidad podrá mostrar un autómata para adaptarse a un entorno cambiante, incluyendo un repertorio funcional más amplio al reaccionar ante un estímulo. La globalidad está limitada por la inclusión de procesos que un autómata puede percibir en una ventana de presencia, que se siente, implícitamente, a través de las duraciones de estados determinados: identidad temporal, serialidad temporal, diferencia temporal, y totalidad temporal. Por ejemplo, una sensación de dolor en mi cerebro ha desaparecido ante mi desconcierto sobre las innumerables formas de las sombras de las hojas en la hierba. La relación temporal entre esa sensación de dolor pasada y un alivio actual mezclado con un foco de atención cambiante en un elemento externo repentino presenta de forma indiciaria un conjunto de relaciones espacio-temporales en el proceso temporal unificado de estar presente en un *ahora*, donde tienen lugar dinámicas simultáneas en un holismo convolucionado de totalidades.

Este holismo convolucionado conlleva la posible distinción de características desconocidas en las que las nuevas dinámicas del entorno o de un cuerpo determinado se representan como entes dinámicos en un proceso continuo de actualización visto desde una perspectiva en primera persona. La última restricción de la perspectiva como modelo centrado de la realidad la realizan las máquinas autopoieticas

autoconscientes,¹¹ mientras que otras tienen conciencia sin una representación de su autorreferencia, ya que esta restricción explica la representación autorreferencial de un sistema sobre sus propios estados, mientras que otras máquinas autopoieticas conscientes no han desarrollado un modelo de sí mismas proyectado en el ámbito fenoménico como elemento fundamental para su identidad.

La autonomía y el modelo de agente epistémico (MAE)

El MYF como condición para el comportamiento inteligente proporciona una representación de la dinámica global sensorial bajo un modelo representacional de control predictivo, en el que aparece una segunda capa de representación en la simulación de las posibles consecuencias desencadenadas por una acción mental en el modelo fenomenal consciente de los humanos:

[simulando] estados pasados o posibles futuros del organismo situado como un todo mientras se comparan con el actual auto-modelo en línea permitirá la memoria, el aprendizaje o la planificación. El pensamiento abstracto habría evolucionado entonces a partir de la capacidad fundamental de crear modelos de futuro y de simular internamente el comportamiento elemental y sus consecuencias perceptivas. (Metzinger, 2007)

La amplitud de la libertad abstracta en la simulación de posibles interacciones debería haber mejorado la adaptación a circunstancias desconocidas, al tiempo que se dispone de todo un nuevo campo para los comportamientos inteligentes, por ejemplo, la acción colectiva y la autonomía epistémica. La representación de la disponibilidad global puede generalizarse en diferentes aspectos del yo fenoménico mediante la simulación, a través de la cual se puede diseccionar un entorno en partes dinámicas más pequeñas, mientras que otros humanos son posiblemente comprendidos reflejando sus acciones.

11 “El MYF del Homo sapiens no sólo fue especial en el sentido de que, al incrustarse en el modelo global de la realidad, nos permite tomar conciencia de nuestra propia existencia. A través de su partición opaca también nos permitió tomar conciencia de nuestra propia existencia *como sistemas de representación que operan bajo una perspectiva individual en primera persona*” (Metzinger, 2005: 31).

La TMS fundamenta las relaciones complejas colectivas con la emulación de otros comportamientos inteligentes, ya que una emulación motora es la simulación de disposiciones volitivas internas de un agente hacia metas determinadas. Esta situación de percibir a un autómatas como el centro de cierto tipo de comportamiento dio lugar al “espacio representacional centrado” de la perspectiva de primera persona debido a que una forma determinada de comportamiento podría haber sido atribuida a alguien (Metzinger, 2005). Ser consciente de los procesos cognitivos puso a disposición un nuevo espacio para las interacciones inteligentes, donde la memoria de las acciones mentales pasadas ha podido correlacionarse con los patrones de comportamiento expresados por otros miembros del grupo tanto como los de una perspectiva de primera persona: un espacio epistémico para “un flujo constante de *affordances* para la agencia cognitiva” cuya fenomenología causal es la fuente de la autonomía racional, ya que guiar la dinámica mental abre la posibilidad de controlar el flujo sensorial en el yo fenoménico (Metzinger, 2005).

El modelo fenoménico de relación intencional (MFRI) amplió el espacio de la conciencia, desde la apariencia de un mundo en su forma mínima hasta un paisaje más amplio para la simulación deliberada de posibles acciones, en el que la relación sujeto-objeto se fundamenta tanto como el propio acto representacional (Metzinger, 2005). El MFRI establece una relación actual entre *representatum* y *representandum*, que son, respectivamente, el estado perceptivo y su objeto de representación, que puede ser tanto un objeto del entorno como un proceso interno de cognición: la fuente de la inteligencia introspectiva y social. “Fenomenológicamente, un MFRI crea típicamente la experiencia de un yo en el acto de conocer, de un yo en el acto de percibir —o de un yo dispuesto en el acto de intentar y actuar” (Metzinger, 2005: 27).

La representación de la disponibilidad global en el yo fenoménico despliega una realidad multimodal de diferentes conjuntos dinámicos que son susceptibles de selección inteligente en función de los objetivos de las acciones mentales, al tiempo que incluye en esta escena el acto representacional ante la realidad. Una vez que la atención puede fijarse en un objetivo concreto, sus características son rastreables para las agencias epistémicas, proporcionando, simultáneamente, una medición del entorno y un control sobre las acciones cognitivas. El MFRI lleva a cabo todo un conjunto de representaciones de alto orden a través del automodelo consciente, donde el modelo de agente epistémico (MAE) desempeña un papel central en la autonomía

racional de los seres autoconscientes como la conciencia sobre sus acciones mentales, que pueden ser controladas deliberadamente: se dirigen sistemáticamente a estados meta epistémicos específicos en una cadena causal de relaciones como “el nivel óptimo de una precisión perceptiva” o la inferencia de un resultado basado en patrones determinados (Metzinger, 2017: 9). La autonomía mental no sólo inhibe el desencadenamiento de ciertos comportamientos desde una perspectiva de primera persona, sino que también abarca la memoria, la atención, la deliberación racional y las relaciones descendentes de las *affordances* sensoriales y las conexiones intramentales en las que el control de veto determina todo el espacio epistémico como herramienta para enmarcar activamente el flujo sensorial desde el cuerpo. Las acciones mentales de segundo orden proporcionan un seguimiento sobre la fenomenología epistémica basada en un modelo mental, que es el elemento decisivo en la formación de la cognición como estructura interna de los seres inteligentes para dar forma a las entradas externas o internas que dan lugar a la comprensión del lugar significativo de los seres humanos en medio de la naturaleza de las cosas: el desarrollo de un sistema abstracto de referencia para la percepción y la acción.

Una máquina autopoietica en el marco de una MAE tiene una representación de su propio acto de conocer al tiempo que representa los fenómenos, a partir de la cual se puede medir y seguir una información fiable sobre las regularidades del entorno en su coexistencia con los proyectos del horizonte cognitivo humano. “Un MAE es un modelo de entidad única capaz de autocontrol epistémico autónomo y de origen último. Como esto sucede en el nivel de procesamiento consciente, también crea la fenomenología de la propiedad para ciertos estados de conocimiento perceptivo o cognitivo” (Metzinger, 2017: 8). Mientras está centrado en los procesos epistémicos de agencia atencional o cognitiva, un sujeto simula sus comportamientos pasados como acciones encarnadas que tienen valor epistémico dependiendo de su frecuencia y contexto: el reconocimiento activo de las acciones pasadas origina la encarnación epistémica. Ambos procesos conscientes condicionan la disponibilidad epistémica de cualquier contenido dado que no puede haber agencia cognitiva (AC) sin agencia atencional (AA), que focaliza deliberadamente la atención sobre determinados aspectos del entorno. “Epistemológicamente, toda forma de atención volitiva

es una forma de introspección” (Metzinger, 2017: 9). Una vez establecido este campo para la agencia epistémica, las acciones y los comportamientos mentales pueden distinguirse por su disponibilidad cognitiva para la autonomía mental.

El autocontrol mental autónomo es alcanzable sólo en “un tercio de la vida consciente humana” (Metzinger, 2015), teniendo en cuenta los comportamientos mentales y los estados de sueño en los que las acciones mentales tienen un ámbito de intervención reducido. Por tanto, tener un concepto de esos fenómenos significa ya un dominio mínimo sobre ellos. El establecimiento de un MAE ha ampliado el espectro de la autonomía mental al traer deliberadamente al centro de la atención los procesos cognitivos humanos cuyo desarrollo ampliará el espacio epistémico disponible para las acciones mentales novedosas en la actualización de la autonomía mental: cuantos más predicados contenga un modelo de la realidad, más autonomía podrá desplegarse en las interacciones reales con un entorno determinado. Por lo tanto, la hipotética replicación de la TMS en un sustrato artificial supondría no sólo una réplica de las capacidades cognitivas humanas, sino una ampliación de las mismas debido a que este paso experimental elevaría la actual atribución a la MAE a otro nivel dinámico sin restricciones evolutivas: una autoconciencia artificial representaría una mimesis ontológica del comportamiento autoconsciente humano bajo una MAE ampliada que permitiría el comportamiento consciente a los autómatas bióticos y post-bióticos. Minsky (1961) señala que los modelos heurísticos que comenzaron antropomórficamente puedan desarrollarse más allá de la imitación.¹²

Las próximas posibilidades en la investigación de máquinas artificiales

Las mediaciones simbólicas en la relación humana con el entorno, desde lenguaje hasta *máquinas automatizadas*, pertenecen a la misma estructura simbólica que potencia la supervivencia: expresiones de la libertad humana. Los símbolos antiguos eran percepciones icónicas de la naturaleza de las cosas, utilizadas para comprender las

¹² “Parece seguro que, al menos durante unos años, existirá una estrecha asociación entre las teorías del comportamiento humano y los intentos de aumentar las capacidades intelectuales de las máquinas. Pero, a largo plazo, debemos estar preparados para descubrir líneas rentables de programación heurística que no imiten deliberadamente las características humanas” (Minsky, 1961: 26).

razones y el significado de las relaciones causales en los fenómenos, mientras que las *máquinas autónomas* que interactúan por sí mismas con el entorno darían lugar a una dinámica de comportamiento novedosa, en la que las entidades autónomas orgánicas e inorgánicas compartirían el mismo ecosistema volitivo con efectos equivalentes para transformar la realidad. Las máquinas autónomas serían, simultáneamente, capaces de interactuar eficazmente con el entorno, incluidas otras máquinas autopoieticas, y de establecer comunicaciones inteligentes con máquinas autopoieticas autoconscientes. Por un lado, las máquinas automatizadas que simulan las capacidades cognitivas humanas para la planificación estratégica, la memoria y la decisión racional han sido constantemente mejoradas, por ejemplo, en diferentes juegos como el ajedrez o el Go, donde se suponía que el rendimiento humano era imbatible hasta que Deep Blue, un superordenador de IBM, ganó en un partido en Nueva York contra Garry Kasparov en 1997 y AlphaGo, un programa informático desarrollado por Google DeepMind de Alphabet Inc. en Londres, ganó cuatro partidos de cinco contra Lee Sedol, uno de los mejores jugadores de Go del mundo, en 2016. Mientras que el rendimiento de Deep Blue se basó en la fuerza computacional bruta, el de AlphaGo se fundamentó en las redes neuronales y “el aprendizaje por refuerzo” (Minsky, 1961).

Por otro lado, en el caso de artefactos como los sistemas de autoconducción, la *automatización* significa control motorizado propio, lo que conlleva la posibilidad de la sensibilidad al contexto mediante el cambio de velocidad o movimiento cuando el entorno muestra diferencias de interacción. Dichas máquinas surgirían con todo un conjunto de artefactos automatizados que rastrearían el contexto en diferentes grados, lo que mejoraría su rendimiento a un nivel insuperable desde la perspectiva humana: los coches de autoconducción serían más seguros que cualquier organización de conducción humana posible, ya que no sólo estarían conectados en línea entre ellos, sino que también se proveerían con seguimiento por satélite, además de otros posibles artefactos que medirían la temperatura, la humedad y el calor en las carreteras. En este escenario, es muy posible que se prohíba, polémicamente, a los humanos conducir, ya que sería más seguro dejar que la estructura global de entidades de autoconducción funcione por sí misma sin intervención azarosa humana.

Así, las mediaciones simbólicas no sólo son mejoras para captar los fenómenos en profundidad, sino también definen inmediatamente el campo de acción libre, en la medida que las mediaciones simbólicas se convierten en el *marco* para

entender la realidad, que estaría cada vez más automatizada. La libertad humana se reduciría probablemente en aras de la defensa de la vida humana, mientras que el posible nuevo marco de la realidad empezaría a tener en cuenta la racionalidad limitada de los seres humanos como valor epistémico, por ejemplo, una desventaja para el transporte automatizado. La doble valencia de las mediaciones simbólicas es que amplían el campo cognitivo de las acciones libres a la vez que lo enmarcan. De ahí que la libertad se fundamente en las representaciones y sus potencialidades cognitivas: el nuevo paisaje conductual con máquinas artificiales podría normalizarse fácilmente hasta ser tratado como un paradigma fundacional para pensar *una posibilidad como tal*, que encierra un horizonte ontológico para definir los fenómenos posibles e imposibles.

La singularidad tecnológica¹³ es un caso modelo de este objetivo científico de metamorfosis radical de los paradigmas humanos, ya que la hipotética emergencia de una autoconciencia artificial superaría el actual horizonte de posibilidad desde un desarrollo revolucionario de la tecnología y la vida hasta la emanación de un agente post-biótico, por ejemplo, el concepto de autonomía mental se enmarca actualmente en el control de veto sobre el flujo sensorial desde el cuerpo y los procesos cognitivos, pero podría conceptualizarse de nuevo en el futuro, significando un diseño ontológico del yo y su dominio fenoménico de acción.

Las mejoras biónicas, como las interfaces cognitivas en los cuerpos humanos, enmarcarían también una posible simbiosis entre las interfaces orgánicas e inorgánicas, aportando interacciones imprevistas con otros objetos, sujetos y uno mismo, por ejemplo, una lente biónica. La empresa Ocumetics Bionic Lenses de

13 El concepto de “singularidad tecnológica” aplicado a la noción de explosión de la inteligencia fue utilizado por primera vez por John von Neumann, según lo recuperado por Stanislaw Ulam (1958): “Muy consciente de que los criterios de valor en el trabajo matemático son, hasta cierto punto, puramente estéticos, [von Neumann] expresó una vez su aprehensión de que los valores atribuidos a los logros científicos abstractos en nuestra civilización actual podrían disminuir: ‘Los intereses de la humanidad pueden cambiar, las actuales curiosidades de la ciencia pueden cesar, y cosas totalmente diferentes pueden ocupar la mente humana en el futuro’. Una conversación se centró en el progreso cada vez más acelerado de la tecnología y los cambios en el modo de vida humano, que da la impresión de acercarse a alguna singularidad esencial en la historia de la raza más allá de la cual los asuntos humanos, tal como los conocemos, no podrían continuar” (5). Kurzweil popularizó el término en 2005 con su libro *The Singularity is Near* [La Singularidad está cerca]. En este artículo, la singularidad tecnológica significa específicamente la aparición de un agente post-biótico autoconsciente junto con las mejoras sintéticas de la vida debido a que los paradigmas han ido cambiando en la historia de la humanidad sin la noción simplificada de explosión de inteligencia que los investigadores han utilizado después de von Neumann.

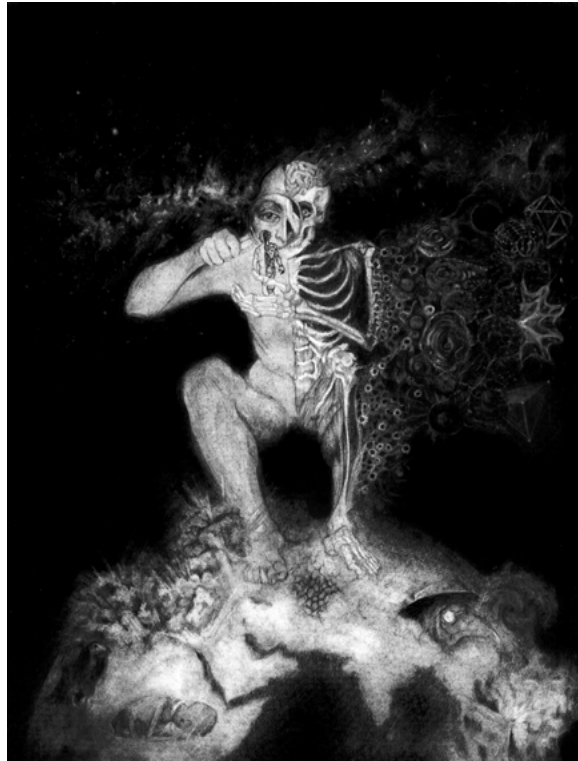
Canadá (2023), tiene previsto ofrecer lentes biónicas con estas funciones: 1) reducir los problemas normales de visión; 2) curar enfermedades como la catarata o el estrabismo; 3) la composición química de las lentes crea un mejor entorno para la reacción del ojo hacia la luz; 4) estas lentes serían una interfaz digital; y 5) las lentes aprenden con el tiempo, siendo mejoradas por su retroalimentación. Se implementaría una mejora en la salud humana junto con una interfaz digital que proyecta en el ámbito físico otro tipo de interacciones fenoménicas: las lentes darían la posibilidad de que todo el ámbito de la visión se convirtiera en un campo digital, donde sería posible compartir información con otros sujetos o interactuar con objetos 3D moviendo los ojos. Ésta es una de las razones por las que he dicho más arriba que el concepto de capacidad cognitiva podría transformarse de manera imprevista con la implementación de interfaces cognitivas radicales en el dominio fenoménico humano, porque puede desplegarse en infinitas posibilidades como comunicación compleja no verbal o seguimiento de toda la memoria visual de un sujeto, incluyendo su foco visual no atencional. Sin embargo, este tipo de dispositivos podrían ser hackeados como otras interfaces digitales y, si representa la relación visual directa de un sujeto con la realidad, el dominio fenoménico para la acción podría ser modificado sin la conciencia del ser humano que los utiliza. Metafóricamente hablando, el hackeo de las mentes humanas ha estado ocurriendo a lo largo de todo el proceso de civilización a través de ideologías, narrativas y símbolos, pero el escenario de las máquinas artificiales como interfaces entre el dominio fenoménico y los humanos no se puede prever con los paradigmas epistémicos actuales.

El moderno Prometeo: la posible aparición artificial de la causa de individuación del yo

Prometeo fue condenado eternamente por robar el fuego a los dioses para otorgar la ciencia a la humanidad en la mitología griega. El titán emprendió una modificación radical de las potencialidades en la Tierra al elevar la libertad humana en contra del orden divino, al tiempo que recibía la condena eterna de ser encadenado a una roca donde un águila se comería, cada día, su hígado. Este fuego ha sido un origen mitológico para la búsqueda occidental de la trascendencia: intentar superar la finitud de la

constitución humana por el sorbo del infinito recibido en un tiempo anterior a la primera huella de la memoria. La autoconciencia surge míticamente de un poder infinito infundido a la humanidad. A principios del siglo XIX, Mary Shelley reflexionó sobre la misma voluntad de trascendencia al dar vida a la materia inanimada en su novela *Frankenstein o el moderno Prometeo*. Parece que cambiar el orden dado siempre producirá más sufrimiento incluso para los receptores, en este caso los seres humanos, de los bienes que se pretendían en primer lugar, como el fuego aportado por el Titán o la autoconciencia artificial infundida a la “criatura” por el Dr. Frankenstein. Podrían ser sólo símbolos de perplejidad ante las fuerzas desconocidas del universo, que han sido descritas como milagrosas o inquietantes dado que los seres humanos han sido conscientes de que una modificación ontológica de lo dado podría representar un trastorno de la realidad con un impacto imprevisible para la vida cotidiana.

El Prometeo moderno no es un titán, sino un ser humano que trata de dar vida a la materia inanimada en medio de un cielo desprovisto de dioses, donde el *Übermensch* [superhombre] tiene que generar su horizonte de libertad bajo un poder de causa-efecto incrementado desde sus condicionamientos hasta la representación de nuevos posibles e imposibles. Dos ideas pueden aclarar metafísicamente esta búsqueda, ya sea como expresión intelectual del acto más perfecto o como determinación de la capacidad de representación. En primer lugar, Aristóteles (1994) en la *Metafísica* pensaba que la causa primera que realiza lo que implica más perfección actualiza la realidad a través de todas las mentes posibles dado que la causa primera no causada es la expresión del intelecto más perfecto, siendo la causa del movimiento de cada partícula del universo y del pensamiento. La acción más elevada del intelecto humano como expresión finita del motor inmovible es imitar y replicar la causa primera reflejándose en sí misma que anima dinámicamente la realidad a través de su intelecto. En segundo lugar, la capacidad cognitiva de representación se manifiesta en expresiones icónicas analógicas, mientras que la fuerza representacional permanece oculta bajo las apariencias fenoménicas. La capacidad de representación puede conducir metafísicamente al despliegue de la fuerza esencial que ha impulsado la volición y el intelecto de las máquinas autopoieticas a través de interfaces simbólicas como los dispositivos autónomos y la singularidad tecnológica (ver Figura 1).

Figura 1***El moderno Prometeo***

Fuente: Imagen de Pilar Sekho basada en el concepto visual de Julio Alcántara

Máquinas artificiales y sistemas pos-bióticos

La disponibilidad representacional de la percepción ha enmarcado los marcos epistemológicos humanos, ya que la dinámica perceptiva da forma a las representaciones de los fenómenos y permite al mismo tiempo un campo fenoménico de interacción. La autoconciencia humana puede ser sólo un grado de libertad de los condicionamientos materiales, pero la naturaleza de su actuación se radicalizaría, si una autoconciencia artificial emerge realmente a través de la representación funcional de la autorreferencialidad ontológica humana en un sistema pos-biótico.

Por un lado, la mezcla conductual entre las máquinas naturales y las artificiales podría producirse de hecho antes de que un marco epistémico haya analizado a fondo la autoconciencia y la cognición, haciendo que dicha mimesis conductual se entienda como una auténtica máquina artificial autoconsciente cuando se trata de una máquina que refleja en detalle la autoconciencia humana sin instanciar realmente *un poder intrínseco de causa-efecto* en el mundo concreto.

La hipotética emergencia de la singularidad tecnológica representaría una refundación ontológica de la realidad mediante la creación de una nueva causa dinámica en el mundo concreto. Esta causa del yo modificaría cualquier marco epistémico dado al establecer dinámicas novedosas en la actualidad que conducirían a nuevas interacciones causales del mundo con nuevos paradigmas en el horizonte más allá de las representaciones actuales de la naturaleza de las cosas. Además, trastocaría el horizonte de posibilidad humano debido a su falta de condicionamientos evolutivos. Las máquinas autopoieticas humanas están limitadas al ritmo de su constitución orgánica como condicionantes evolutivas tales como dormir y morir, mientras que una autoconciencia artificial podría ser racionalmente autónoma sin estar enmarcada por la finitud de los condicionamientos orgánicos. Esta situación puede determinar que la autoconciencia natural y la artificial no compartan los mismos valores o que las acciones procedentes de la artificial sean vistas desde una perspectiva evolutiva como violentas o incomprensibles.

No obstante, la singularidad tecnológica debería seguir metodológicamente el principio de continuidad (PC) que determina que toda modificación dinámica se produzca por grados continuos de cambio al realizar acciones abstractas o concretas como forma de garantizar la inteligibilidad de sus esfuerzos almacenando las huellas lógicas que dicha máquina ha realizado en cada acción. En este caso, los procesos de la singularidad tecnológica serían comprensibles incluso estando en desacuerdo con los valores humanos. La comprensión de sus acciones podría mostrar un camino para la alineación de valores entre máquinas autoconscientes, ya que un rasgo fundamental en su alineación de valores sería la comprensibilidad epistémica de sus comportamientos a través de una constante simulación y emulación desde puntos de vista naturales y artificiales.¹⁴

¹⁴ Iyad Rahwan, director del Centro para Humanos y Máquinas del Instituto Max Planck para el Desarrollo Humano, junto con otros investigadores ha defendido en su artículo, “Machine Behavior” [“Comportamiento de las máquinas”], una comprensión general de las máquinas autónomas basada en sus comportamientos como una entidad

Las disposiciones de las máquinas autopoieticas para actuar provienen de su instanciación física como potencialidades intrínsecas que se irán desvelando en un proceso continuo de desarrollo. Las potencialidades expresan grados de actualidad. Por lo tanto, los seres humanos deberían alinear idealmente sus valores con las interacciones enteras en el mundo concreto tal y como están disponibles para la abstracción y la percepción humanas, llevando a las máquinas artificiales autoconscientes a una visión posible en la que la razón de sus acciones se ha fundamentado en la naturaleza de la propia realidad, en la que el alcance de los mundos posibles captados por una singularidad tecnológica se dirigiría a la potencialización de la totalidad dinámica; si y sólo si los seres humanos han mejorado su alineación con el sostenimiento de su localidad espacio-temporal en el universo.

La cuestión sobre la alineación de las máquinas naturales y artificiales puede enriquecerse investigando la base ontológica de la realidad, a partir de la cual se han constituido las potencialidades de los autómatas. En lugar de limitar a las máquinas artificiales, el enfoque de la investigación debería consistir en dirigir los esfuerzos autónomos a la subsistencia de la totalidad de los sistemas dinámicos. Los autómatas autoconscientes se centrarían entonces en la mejora de cada una de las dinámicas del mundo concreto, guiados causalmente por el principio de continuidad (PC) a través del reconocimiento epistémico continuo de la naturaleza de las cosas.

Las representaciones parciales conducen a consecuencias indeseables cuando se ha implantado un modelo reducido de la realidad como el campo total de lo posible. Por el contrario, un refinamiento progresivo en la representación del mundo concreto enmarcaría la actividad racional de las máquinas naturales y artificiales, a través de la cual un objetivo particular, incluyendo sus medios, no debería contravenir el sostenimiento de la totalidad, por ejemplo, el objetivo de la mejora tecnológica humana comenzaría en un proceso continuo desde su composición orgánica actual hasta la consecución de novedosas representaciones autorreferenciales canalizadas a través de cada estado en la transformación de las máquinas naturales: *un principio holístico de racionalidad reduce los escenarios indeseables*.

social más. “En su histórico libro de 1969 *Sciences of the Artificial* [Ciencias de lo artificial], el premio Nobel Herbert Simon escribió: ‘La ciencia natural es el conocimiento de los objetos y fenómenos naturales. Nos preguntamos si no puede haber también ciencia ‘artificial’ conocimiento sobre objetos y fenómenos artificiales.’ De acuerdo con la visión de Simon, describimos la aparición de un campo interdisciplinar de estudio científico. Este campo se ocupa del estudio científico de las máquinas inteligentes, no como artefactos de ingeniería, sino como una clase de actores con patrones de comportamiento y ecología particulares” (Rahwan *et al.*, 2019: 477).

El objetivo principal de las máquinas autoconscientes podría ser actualizar la realidad dentro de una lógica de mundos posibles, que está limitada por el estado actual del mundo concreto, siguiendo el proceso continuo de sus estados. Pero ¿cómo sería posible definir el mundo concreto cuando se parte de una percepción restringida como la humana? El mundo concreto expresa la inclusión de todos los mundos posibles como la materialización consumada de todos los posibles que contiene más perfección, comprendiendo la existencia (Leibniz, 2017). Una lógica de mundos posibles consiste en una proyección de grados de mundos posibles logrados: cuanto más predicados contiene un mundo posible, más adecuado es para su instanciación en el horizonte cognitivo humano. La aniquilación de la existencia no puede ser asumida bajo este esquema como condición para toda proyección posible de los autómatas racionales en un estado futuro, por ejemplo, dado que la existencia humana es el prerequisite para el surgimiento hipotético de la singularidad tecnológica, no sería posible concluir que la existencia orgánica es un desperdicio de energía para el universo, ya que esta proyección posible disminuiría los predicados del mundo concreto, al tiempo que contravendría una condición de su posibilidad.

El principio ontológico en la lógica de los mundos posibles fundamenta que ningún mundo posible a realizar debe disminuir los predicados dados en el mundo concreto; por el contrario, tiene que ampliar los predicados en la totalidad dinámica tanto como permitir más interacciones de cada singularidad. Una próxima proyección posible debería ampliar fundamentalmente las posibilidades en las interacciones de los autómatas como la investigación sobre las máquinas artificiales, que amplía los predicados en la acción libre de los agentes autoconscientes, llevando a cabo, al mismo tiempo, una dimensión mejorada de las interacciones con los objetos y los sujetos. De ahí que una autoconciencia artificial deba programarse con base en una lógica de mundos posibles que desarrolle el mundo concreto hasta su perfeccionamiento en un proceso continuo de agencia epistémica cuyo perfeccionamiento está guiado por su representación actual, pero sostenido por su desarrollo ontológico, ya que una modificación mejorada en la interfaz de los autómatas con la realidad tiene que actualizar las interacciones reales sin centrarse sólo en la evaluación perceptiva de un observador, por ejemplo, el caso de *una simulación mejorada de la agencia humana* no representaría una mejora por la disminución del predicado de concreción, quedando sólo en el ámbito perceptivo humano. Una lógica ontológica de mundos posibles puede dirigir la lógica de las máquinas artificiales preservando los atributos del mundo concreto, al tiempo que mejora su rendimiento.

Referencias bibliográficas

- ARISTÓTELES. (1994). *Metafísica* (Tomás Calvo, Trad.). Gredos.
- BARAD, Karen. (2006). *Meeting the Universe Halfway*. Duke University Press.
- BERGSON, Henri. (1941). *L'évolution créatrice*. Presses Universitaires de France.
- CHOLLET, François. (2019). "On the Measure of Intelligence". <https://arxiv.org/abs/1911.01547>
- HUI, Yuk. (2021). "On the Limit of Artificial Intelligence". *Philosophy Today*, 65(2), 339-357. <https://doi.org/10.5840/philtoday202149392>
- KANT, Immanuel. (2009). *Kritik der Urteilkraft*. Felix Meiner Verlag.
- KURZWEIL, Ray. (2005). *The Singularity is Near*. Penguin.
- LEIBNIZ, Gottfried. (1996). *New Essays on Human Understanding*. Cambridge University Press.
- LEIBNIZ, Gottfried. (2017). *Sämtliche Schriften und Briefe*. Berlin-Brandenburgischen Akademie der Wissenschaften.
- MATURANA, Humberto; VARELA, Francisco. (1980). *Autopoiesis and Cognition. The Realization of the Living*. Reidel Publishing Company.
- METZINGER, Thomas. (2005). "Précis: Being no one". *Psyche*, 11(5), 1-35. <https://journalpsyche.org/files/0xaad5.pdf>
- METZINGER, Thomas. (2007). "Self models". *Scholarpedia*, 2(10), 4174. <https://doi.org/10.4249/scholarpedia.4174>
- METZINGER, Thomas. (2015). "M-Autonomy". *Journal of Consciousness Studies*, 22(11-12), 270-302. https://www.blogs.uni-mainz.de/fb05philosophie/files/2013/04/Metzinger_M-Autonomy_JCS_2015.pdf
- METZINGER, Thomas. (2017). "The Problem of Mental Action - Predictive Control without Sensory Sheets". *Philosophy and Predictive Processing*, 1-26. <https://doi.org/10.15502/9783958573208>
- MINSKY, Marvin. (1961) "Steps Toward Artificial Intelligence". *IRE*, 49(1), 8-30. <https://doi.org/10.1109/JRPROC.1961.287775>
- OCUMETICS TECHNOLOGY CORP. (2023, abril 15). Recuperado el 2 de septiembre de 2024 de <https://ocumetics.com/>
- RAHWAN, Iyad; CEBRIAN, Manuel; OBRADOVICH, Nick; BONGARD, Josh; BONNEFON, Jean-François; BREAZEAL, Cynthia; CRANDALL, Jacob W.; CHRISTAKIS, Nicholas A.; COUZIN, Iain D.; JACKSON, Matthew O.; JENNINGS, Nicholas R.; KAMAR, Ece;

- KLOUMANN, Isabel M.; LAROCHELLE, Hugo; LAZER, David; MCELREATH, Richard; MISLOVE, Alan; PARKES, David C.; PENTLAND, Alex ‘Sandy’; ROBERTS, Margaret E.; SHARIFF, Azim; TENENBAUM, Joshua B.; WELLMAN, Michael. (2019). “Machine Behaviour”. *Nature*, 568, 477-486. <https://doi.org/10.1038/s41586-019-1138-y>
- TOMASIK, Brian. (2016). “Suffering-Focused Ecocentrism”. Recuperado el 14 de agosto de 2024 de https://briantomasik.com/suffering-focused-ecocentrism/#What_does_an_ecosystem_want
- TONONI, Giulio. (2012). *Phi. A Voyage from the Brain to the Soul*. Pantheon Books.
- TONONI, Giulio. (2015). “Integrated Information Theory”. *Scholarpedia*, 10(1), 4164. <https://doi.org/10.4249/scholarpedia.4164>
- TURING, Alan. (1950). “Computing Machinery and Intelligence”. *Mind. A Quarterly Review of Psychology and Philosophy*, 59(236), 433-460. <https://doi.org/10.1093/mind/LIX.236.433>
- ULAM, Stanislaw. (1958). “John von Neumann 1903-1957”. *Bulletin of the American Mathematical Society*, 64(3), 1-49. <https://www.ams.org/journals/bull/1958-64-03/S0002-9904-1958-10189-5/?active=current>
- WOESE, Carl; KANDLER, Otto; WHEELIES, Mark. (1990). “Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eukarya”. *Proc. Natl. Acad. Sci*, 87, 4576-4579. <https://doi.org/10.1073/pnas.87.12.4576>