COLLOQUIA

*Revista de Pensamiento y Cultura Vol. **12** (2025), págs. 158 - 183. ISSN: 1390-8731*
DOI 10.31207/colloquia.v12i1.196

Un giro copernicano para la IA: De la ética informacional a la Responsabilidad Ontogénica

A Copernican shift for AI: From informational ethics to ontogenic responsibility

Raúl López Ramírez
raul.lopez.r@ug.uchile.cl
Universidad de Chile
ORCID ID 0009-0000-6904-187X

Resumen: El presente trabajo propone un "giro copernicano" para la Inteligencia Artificial (IA), desplazando el centro ontológico desde el paradigma cognitivista dominante hacia el enfoque enactivo, lo que genera cambios radicales en su comprensión ética. Se argumenta que la concepción de la cognición como procesamiento de información, cuya expresión paradigmática en filosofía de la información es el sistema de Luciano Floridi, opera como un "universo ptolemaico" que se basa en un error categorial: tratar la información como una entidad preexistente —un "flogisto moderno"—. Desde la alternativa enactiva, centrada en la autopoiesis y el acoplamiento estructural, se sostiene que el éxito de la IA, más allá de validar el modelo cognitivista, puede interpretarse como evidencia paradójica de nuestra profunda plasticidad estructural. Esta posición permite proponer un fundamento ético de mayor alcance, articulado sobre la noción de "Principio de Responsabilidad Ontogénica", enfocado no en la gestión del artefacto, sino en la co-evolución consciente de nuestro modo de ser con la tecnósfera que habitamos.

Palabras clave: Ética de la IA; enactivismo; autopoiesis; responsabilidad ontogénica; cognitivismo.

Recepción: 01-10-2025 Aceptación: 08-12-2025 Publicación electrónica: 20-12-2025

Abstract: This paper proposes a "Copernican turn" for Artificial Intelligence (AI), shifting the ontological center from the dominant cognitivist paradigm to the enactive approach, a move that radically alters the ethical understanding of it. It is argued that the conception of cognition as information processing, whose paradigmatic expression in philosophy of information is Luciano Floridi"s system, operates as a "Ptolemaic universe" based on a category error: treating information as a pre-existing entity—a "modern phlogiston". From the enactive alternative, centered on autopoiesis and structural coupling, it is maintained that the success of AI, rather than validating the cognitivist model, can be interpreted as paradoxical evidence of our profound ontogenetic plasticity. This stance allows for the proposal of a more far-reaching ethical foundation, articulated around the "Principle of Ontogenic Responsibility"—one focused not on managing the artifact, but on the conscious co-evolution of our being with the technosphere we inhabit.

Keywords: AI ethics; enactivism; autopoiesis; ontogenic responsibility; cognitivism.

Introducción

En la historia de la ciencia, pocas revoluciones han sido tan profundas como el giro copernicano (Kuhn, 1993). Durante siglos, el modelo de Ptolomeo —con la Tierra inmóvil en el centro del universo— dominó el pensamiento astronómico. Era un sistema ingenioso y predictivamente útil, pero para dar cuenta del movimiento anómalo de los planetas requería de una maquinaria conceptual cada vez más compleja de epiciclos y deferentes. El giro propuesto por Copérnico fue, en esencia, un acto de simplicidad radical: al postular que la Tierra no era el centro del universo, sino un planeta más que orbitaba alrededor del Sol, no solo se ofrecía una mejor explicación, sino que se transformaba por completo la comprensión del cosmos y nuestro lugar en él.

Este artículo sostiene que nuestra comprensión de la inteligencia, y por extensión de la Inteligencia Artificial (IA), puede avanzar hacia un giro

copernicano análogo, un desplazamiento fundamental del centro de gravedad ontológico con profundas implicaciones para nuestra comprensión de la mente y la propia condición humana en la era digital. El paradigma dominante que ha impulsado el desarrollo de la IA durante más de medio siglo puede ser entendido como el universo "ptolemaico" de la cognición. Arraigado en el funcionalismo computacional que nació a partir de las ideas de Turing (1950) y consolidado por el éxito notable del aprendizaje automático, este paradigma concibe la cognición como un proceso de cómputo representacional. En esta visión del mundo, la realidad es un universo externo, objetivo y rico en información, y la mente —ya sea biológica o artificial—, es un "planeta" que orbita a su alrededor cuya función primordial es captar, procesar y generar un modelo interno de dicha información que sea útil para la consecución de los fines del agente. Se trata de una cosmología donde la información es el centro inmóvil y la inteligencia es la capacidad de utilizarla para adecuar la conducta a los desafíos del entorno.

Esta cosmovisión ha encontrado en la obra de Luciano Floridi (2024) a uno de sus "astrónomos" más sistemáticos y sofisticados, quien ha dotado a este universo informacional de una compleja y elegante "mecánica celeste" —su Realismo Estructural Informacional (REI)—, ofreciendo una de las arquitecturas conceptuales más robusta para la era digital. Sin embargo, desde hace décadas, una perspectiva radicalmente distinta ha venido gestándose, una que propone un desplazamiento completo de este centro ontológico. El enfoque enactivo, nacido de la biología del conocimiento de Humberto Maturana (1970) y el enactivismo de Francisco Varela (1990), no busca ajustar las órbitas del modelo ptolemaico, sino proponer un nuevo "sol". En este giro copernicano, el centro del universo cognitivo ya no es el mundo externo y su información, sino el agente vivo y autónomo. La tesis enactivista clásica sostiene que un organismo no procesa un mundo pre-dado, sino que, a través de su acción corporeizada y su acoplamiento con el entorno, enactúa o hace emerger un mundo de significado. La información, en esta nueva "cosmología", deja de ser el punto de partida para convertirse en una consecuencia: una distinción que emerge de la danza relacional entre un sistema y su mundo, una danza cuyo único fin intrínseco es la continuación de la vida misma.

El recorrido de este trabajo se articulará en tres fases. En primer lugar, se reconstruirá la arquitectura del "universo ptolemaico", trazando la genealogía del paradigma representacional desde sus raíces filosóficas hasta su consagración en la IA y su culminación en el sofisticado sistema de Luciano Floridi. En segundo lugar, se presentará la alternativa enactiva, mostrando cómo su fundamento en la autopoiesis y el acoplamiento estructural no solo ofrece una nueva cosmovisión, sino que permite interpretar la "información" del paradigma clásico como un "flogisto moderno": un error categorial que, desde esta perspectiva, nos ha impedido comprender la naturaleza profunda del conocimiento. Finalmente, se mostrará que la confrontación entre ambos paradigmas expone un punto ciego fundamental en la ética de la IA contemporánea. Al desplazar el foco desde la mera gestión del artefacto hacia nuestra co-transformación con él, este trabajo culminará con la propuesta de un principio que sirva como nuevo fundamento ético: el Principio de Responsabilidad Ontogénica, una brújula para navegar conscientemente la deriva de nuestro propio ser en el cosmos tecnológico que co-creamos y habitamos.

El paradigma cognitivista: arquitectura del universo "ptolemaico"

Para comprender la magnitud del giro copernicano que el enactivismo propone, es indispensable primero cartografiar el "universo ptolemaico" que viene a desplazar. Este marco conceptual, que ha impulsado el desarrollo de la IA durante más de medio siglo, no es un mero capricho tecnológico, sino la culminación de una larga trayectoria en la filosofía de la mente. Nos referimos al Paradigma Cognitivista, un programa de investigación cuya tesis central es que la cognición es un proceso de cómputo sobre representaciones internas.

El legado funcionalista: la solución al problema mente-cuerpo

La formulación moderna del problema de la inteligencia queda marcada por la distinción dualista de René Descartes (2010) entre la *res cogitans* (la sustancia pensante, inmaterial) y la *res extensa* (la sustancia material). Esta brecha ontológica generó una serie de respuestas materialistas que buscaron

naturalizar la mente, pero fue el funcionalismo la que se consolidó como la solución más influyente y la piedra angular del cognitivismo (Kim, 2011).

El funcionalismo propone que un estado mental no se define por el material del que está hecho, sino por el rol causal o la *función* que desempeña en la economía cognitiva de un sistema. Esta idea, conocida como el principio de realizabilidad múltiple, sostiene que la mente es al cerebro lo que el software es al hardware (Churchland, 1999). La misma función mental (ej. sentir dolor, creer que llueve) podría ser implementada en sustratos tan diversos como un cerebro biológico, un circuito de silicio o, hipotéticamente, cualquier otro sistema con la organización causal adecuada.

Esta abstracción de la mente de su base material fue anticipada por Alan Turing (1950), quien propuso un criterio puramente operacional para la inteligencia, eludiendo el debate ontológico sobre su esencia. Dicha visión encontró su consagración institucional en el proyecto de investigación de Dartmouth de 1955, que acuñó el término "Inteligencia Artificial" y estableció su influyente definición en torno al problema de "hacer que una máquina se comporte de un modo que se llamaría inteligente si un ser humano se comportara de ese modo" (McCarthy et al., 1955, p. 11). Esta definición contrafactual, herencia directa del funcionalismo temprano de Turing, sentó las bases anti-esencialistas sobre las que se construiría todo el paradigma dominante de la IA.

Los pilares del cognitivismo: representación y cómputo

El paradigma cognitivista se erige entonces sobre dos supuestos filosóficos interconectados, que constituyen su núcleo y que el sistema de Floridi lleva a su máxima expresión: el representacionalismo y el computacionalismo.

El primer pilar es el representacionalismo. El funcionalismo, al definir los estados mentales por su rol causal, deja abierta la pregunta sobre qué es aquello que la mente manipula. La respuesta cognitivista es que la mente opera sobre representaciones internas. Esta visión presupone la distinción cartesiana entre un mundo interno y uno externo, objetivo y rico en información, que existe con independencia del observador (Newell & Simon,

1976). La función primordial de la cognición es construir un modelo interno de esa realidad externa que permita al agente navegar el mundo y alcanzar sus fines. En su versión más fuerte, este modelo busca ser un reflejo fidedigno, un "espejo de la naturaleza" (Rorty, 1979), pero su objetivo fundamental es siempre la utilidad para la acción. La mente, por tanto, es un sistema que crea, almacena y transforma representaciones simbólicas del mundo (Fodor, 1981).

El segundo pilar es el computacionalismo, que especifica *cómo* se procesan dichas representaciones. La tesis es que el pensamiento es una forma de cómputo. Las transiciones entre estados mentales se rigen por reglas formales, análogas a los algoritmos de un programa de ordenador, que operan sobre la sintaxis de las representaciones internas. Esta visión da origen al llamado "modelo sándwich" de la cognición (Hurley, 1998), en el cual la percepción (*input*) y la acción (*output*) son meros periféricos de un procesador central y abstracto donde ocurre el "verdadero pensamiento". En síntesis, como lo resume Skidelsky:

> La teoría representacional/computacional de la mente sostiene que los procesos cognitivos son procesos computacionales que operan transformaciones causales sobre los contenidos de estados mentales. Estos contenidos están constituidos por representaciones mentales, entendidas como símbolos que poseen propiedades sintácticas (o físicas) y semánticas (son intencionales, i. e., acerca de algo). (2023, p. 63)

Las olas críticas: de la emergencia conexionista a la ruptura enactiva

El edificio conceptual del cognitivismo, con su arquitectura de símbolos y reglas, dominó el panorama intelectual durante décadas.[1] Sin embargo, sus cimientos comenzaron a mostrar fisuras bajo la presión de interrogantes que su propio marco no podía resolver. La crítica no llegó como un asalto único, sino en dos grandes olas sucesivas, cada una más radical que la anterior. La

[1] Específicamente, durante el periodo que va desde eventos fundacionales como el Taller de Dartmouth de 1956 (cf. McCarthy et al., 1955) hasta el resurgimiento del conexionismo a mediados de la década de 1980 (cf. Rumelhart & McClelland, 1986).

primera cuestionó los componentes centrales de su arquitectura: su unidad de representación —el símbolo— y su mecanismo de procesamiento serial, mientras que la segunda atacó su supuesto más fundamental: la representación.

La primera ola: el conexionismo. La primera gran alternativa al paradigma cognitivista surgió de una insatisfacción con su maquinaria central. La idea de que la inteligencia opera mediante reglas secuenciales aplicadas a símbolos discretos se enfrentaba a dos problemas evidentes: era biológicamente implausible y computacionalmente frágil. El cerebro no opera como un procesador serial —el famoso "cuello de botella von Neumann" (Backus, 1978)— y los sistemas simbólicos son notoriamente quebradizos ante la pérdida de información (Churchland, 1999).

Inspirado en la estructura masivamente paralela del cerebro, el conexionismo —o el paradigma "subsimbólico", como lo denomina Varela (1990)— propuso una arquitectura radicalmente distinta. En esta visión, la cognición no es el producto de una computación simbólica, sino la emergencia de estados globales en una red de componentes simples (redes neuronales) interconectados (Hopfield, 1982). El conocimiento no reside en símbolos localizados, sino que está distribuido en la fuerza de las conexiones de toda la red, y el sistema aprende ajustando dichas conexiones a través de la experiencia (Rumelhart & McClelland, 1986).

Este enfoque representó un avance fundamental. Demostró que conductas cognitivas complejas podían emerger de reglas locales simples, sin necesidad de un procesador central que manipulara representaciones explícitas. Sin embargo, a pesar de revolucionar el mecanismo, el conexionismo en su forma estándar no siempre rompió con el marco cognitivista. A menudo, el objetivo de la red seguía siendo el mismo: resolver un problema o representar de manera óptima un mundo que se asumía como

pre-dado. La crítica había desmantelado el símbolo y su proceso, pero dejaba intacto el pilar de la representación (Varela, Thompson, & Rosch, 1991).[2]

La evolución hacia un criterio de utilidad para la acción puede entenderse como un peldaño crucial en una escalera de críticas. Si el cognitivismo clásico se centraba en la *fidelidad* de la representación, la primera ola crítica desplazó el foco hacia su *utilidad*. Sin embargo, este importante avance pragmático deja una pregunta fundamental sin resolver y sirve de antesala a la ruptura enactiva. El enactivismo, como se verá a continuación, toma en serio este giro hacia la acción, pero lo lleva a su conclusión lógica y radical: si la acción efectiva es el verdadero criterio de la inteligencia, ¿es necesaria en absoluto la mediación de una representación?

La segunda ola: la ruptura enactiva. Esta insatisfacción más profunda, por tanto, dio origen a la segunda y más radical ola crítica. La nueva perspectiva, como explicitan los propios pioneros del enactivismo (Varela, Thompson, & Rosch, 1991), se nutrió de tradiciones como la fenomenología de la percepción de Merleau-Ponty (1993) para atacar el problema de raíz. El foco ya no era *cómo* se representa el mundo, sino la idea misma de que la cognición consiste en representar un mundo pre-dado. Esta crítica apuntaba directamente al residuo cartesiano que el cognitivismo había heredado: el dualismo sujeto-objeto y la concepción de la cognición como un proceso desencarnado.

Partiendo de la biología, la alternativa enactiva se fundamenta en la autopoiesis, la organización que define a los seres vivos como sistemas autónomos que se producen y mantienen a sí mismos (Maturana & Varela, 1973). En sus desarrollos posteriores, para dar cuenta de la normatividad

[2] Es importante subrayar que esta persistencia del marco representacionalista no es una mera anécdota histórica. Los actuales sistemas de aprendizaje profundo, herederos directos de esta primera ola, son un testimonio de esta ambigüedad: aunque su arquitectura subsimbólica es radicalmente distinta a la de la IA clásica, su función primordial sigue siendo, en la mayoría de los casos, optimizar una correspondencia con un vasto conjunto de datos que se asume como un reflejo del mundo. De este modo, aunque el *mecanismo* computacional se ha vuelto más sofisticado y biológicamente plausible, el *supuesto ontológico* de fondo —que la inteligencia opera sobre un mundo pre-dado y rico en información— a menudo permanece incuestionado.

graduada (estados mejores o peores) propia del actuar de los organismos, este concepto se ha complementado con el de adaptatividad, entendida como la capacidad del sistema de regular activamente su relación con el entorno para mantenerse dentro de sus límites de viabilidad.[3] La clausura operacional autopoiética significa que las interacciones de un sistema vivo con el entorno son siempre moduladas por su propia estructura, la cual se conserva. Esta historia ininterrumpida de transformaciones congruentes entre el sistema y su entorno es lo que se denomina acoplamiento estructural (Maturana & Varela, 1987).

Desde esta base, la cognición se redefine por completo. Se convierte en enacción: la "puesta en obra" o el "hacer emerger" un mundo de significado a través de la acción corpórea. La percepción es acción guiada perceptivamente; las estructuras cognitivas emergen de patrones sensorio-motores recurrentes. El mundo y el conocedor se co-determinan y surgen juntos en una historia de acoplamiento (Varela, Thompson, & Rosch, 1991). Parafraseando el verso de Machado (1912), que sirve de metáfora central para este enfoque, la cognición es "una senda que no existe pero que se hace al andar".

Esta ruptura enactiva disuelve los supuestos cognitivistas:

1. Frente a un mundo pre-dado, postula un mundo enactuado que emerge de la historia de acciones del agente.

2. Frente a la cognición como representación, postula la cognición como acción efectiva, donde la inteligencia no se mide por la veracidad de un modelo, sino por la viabilidad de la acción.

3. Frente al modelo sándwich, postula la cognición corpórea, donde percepción, acción y pensamiento son inseparables y constitutivos de un mismo proceso.

[3] Esta evolución conceptual es clave, pues es la conjunción de la autonomía (autopoiesis) y la adaptatividad lo que fundamenta la noción de creación de sentido (sense-making), concepto central para la perspectiva enactiva (cf. Di Paolo, 2005; Thompson, 2007).

4. Frente a una cognición abstracta y desinteresada, postula una cognición afectiva. El mundo que un agente enactúa no es neutral, sino un dominio de significado y valoración, donde lo afectivo es inseparable del conocer. Al poner la relevancia y el cuidado (*concern*) en el centro, el enactivismo reúne la epistemología y la ética, sentando las bases para una comprensión no-dualista de la mente.

Esta segunda ola crítica, por lo tanto, no ofrece una simple actualización del modelo mental. Propone un cambio de paradigma total: un desplazamiento desde la concepción de la mente como un "espejo de la naturaleza" hacia una comprensión de la mente como un proceso incesante de construcción de mundos a través de la acción, la historia y la corporalidad.

La cognición se comprende, por lo tanto, como la acción efectiva y corporeizada mediante la cual un sistema autopoiético mantiene su viabilidad en una historia de acoplamiento estructural (Varela, Thompson, & Rosch, 1991). La inteligencia, bajo esta nueva luz, es la expresión de una necesidad biológica ineludible: la continuación de la vida. Deja de ser una propiedad abstracta para convertirse en "la capacidad de un ser vivo, en cualquier dominio, de tomar parte en una conversación" (Maturana & Guilloff, 2006, p. 15), entendiendo "conversación" en su sentido más amplio como un fluir de coordinaciones conductuales recurrentes. Esta reconceptualización radical de la mente transforma por completo el terreno sobre el que se ha de fundamentar la ética para la era tecnológica en que vivimos.

Antes de proceder, sin embargo, conviene hacer una advertencia metodológica. La tradición autopoiética que hemos revisado como alternativa al cognitivismo no es un bloque monolítico. En su interior alberga un rico debate sobre los fundamentos biológicos del conocimiento y el origen del significado, que aquí solo podemos mencionar. Por un lado, la vertiente del enactivismo clásico, de inspiración fenomenológica, entiende la autonomía del ser vivo como la base para un genuino "hacer emerger un mundo de significado" (*sense-making*), posicionando al organismo como un centro de perspectiva y valoración.

Por otro lado, una lectura mecanicista más estricta de la Teoría Autopoiética cuestiona esta asimetría cognitiva (cf. Villalobos & Silverman, 2018), proponiendo la relación organismo-entorno como una deriva simétrica y acéntrica, donde el "significado" no es una propiedad del fenómeno biológico, sino una atribución que el observador humano realiza en el lenguaje. Explorar a fondo esta bifurcación, que es concebida incluso como inconmensurable (cf. Villalobos & Palacios, 2021), así como otras recientes e innovadoras propuestas de síntesis, como el "enactivismo computacional" bajo el Principio de la Energía Libre (Korbak, 2021), excede los objetivos del presente trabajo. No obstante, en lo que respecta al propósito de esta crítica a la ética informacional, estas vertientes pueden interpretarse como un frente unido, pues todas coinciden en la insuficiencia de un modelo que ignora la condición encarnada y auto-organizadora de la vida. Es este punto de acuerdo fundamental el que permite una propuesta ética de mayor alcance.

Del paradigma cognitivista a la ética de la IA

La culminación cognitivista: el Realismo Estructural Informacional de Floridi

El paradigma cognitivista, con sus pilares funcionalistas y computacionales, encuentra su expresión más sistemática y ambiciosa en la obra de Luciano Floridi (2024). Partiendo de la definición operacional de IA de Dartmouth, Floridi erige un completo sistema filosófico para interpretarla: el Realismo Estructural Informacional (REI). El REI es una forma de Realismo Estructural Óntico que sostiene que la realidad fundamental no está constituida por objetos, sino por estructuras, y el giro distintivo de Floridi es especificar que estas estructuras son de naturaleza informacional (Floridi, 2008).

En su nivel más profundo, la realidad última no es la información semántica, sino los datos pre-semánticos (*dialephora*): diferencias puras, "faltas de uniformidad" que son la materia prima ontológica (Floridi, 2010). Es solo cuando estos datos se organizan y estructuran que se convierten en información. Esta ontología —"it from data"— es más modesta que el Pancomputacionalismo —"it from bit" (Wheeler, 1990)—, pues no afirma

que el universo compute, sino solo que está constituido por datos. Este marco fundamenta su concepto de una infoesfera (Floridi, 2011): el entorno global constituido por la totalidad de los procesos, agentes y propiedades informacionales en el que habitamos una vida híbrida, o según sus términos, "onlife" (Floridi, 2015).

Para analizar esta infoesfera, Floridi emplea una útil herramienta metodológica: los Niveles de Abstracción (NdA) (Floridi & Sanders, 2004). Un NdA es una "interfaz" o punto de vista explícito, compuesto por un conjunto de "observables" (variables tipadas e interpretadas), que permite modelar un sistema. Su función estratégica es crucial: le permite defender al REI de las críticas fenomenológicas, relegando la experiencia vivida a un NdA entre otros, válido en su propio dominio pero no ontológicamente primario. Sin embargo, este método presupone, por definición, la figura de un observador externo y teleológico que elige los observables según sus propios fines —un punto ciego epistemológico que será fundamental para la crítica posterior—.

Armado con este andamiaje, Floridi formula su influyente tesis, que representa la consecuencia lógica y la extremación del principio funcionalista de la realizabilidad múltiple: "la IA no es una nueva forma de inteligencia, sino una nueva forma de agencia" (Floridi, 2024, p. 137). Se trata de una agencia informática, definida funcionalmente por su capacidad de procesar datos, actuar de forma autónoma y aprender de sus interacciones, sin necesidad alguna de estados mentales o conciencia. Lo revolucionario es el "divorcio" que esto implica: la IA ha logrado "*desacoplar* la capacidad de resolver un problema o completar una tarea con éxito de cualquier necesidad de ser inteligente para hacerlo" (Floridi, 2024, p. 62).

El éxito de esta agencia no-inteligente se explica no porque las máquinas se vuelvan más como nosotros, sino porque estamos rediseñando el mundo para que sea más como ellas. Esto se logra mediante dos estrategias clave (Floridi, 2024):

- Envoltura (*Enveloping*): Consiste en rediseñar nuestros entornos para que sean más favorables a la IA, transformando tareas

difíciles (que requieren destreza corporeizada) en tareas complejas (computacionalmente intensivas pero manejables para la IA). El ejemplo paradigmático es el vehículo autónomo: no ponemos un androide en el asiento del conductor, sino que "envolvemos" el ecosistema de transporte con sensores y redes, eliminando la necesidad de la destreza humana.

- Ludificación (*Gamification*): La IA alcanza su máximo potencial cuando un proceso puede ser transformado en un "juego" con reglas constitutivas (como el ajedrez), lo que le permite generar sus propios datos sintéticos y alcanzar un rendimiento sobrehumano.

En definitiva, la propuesta de Floridi consolida una visión de la IA como una agencia puramente funcional y sin propósito intrínseco, cuyo éxito depende de nuestra capacidad para re-ingeniar la realidad a su medida. Es la expresión más sistemática del paradigma cognitivista, y es precisamente esta culminación la que revela las tensiones que serán el foco de la presente crítica.

La información como "flogisto moderno": la crítica radical de Varela

La ontología informacional de Floridi, si bien es la culminación del paradigma cognitivista, se asienta sobre un pilar conceptual que el enfoque enactivo no solo cuestiona, sino que busca superar: la noción misma de "información" como una entidad preexistente. Para articular la profundidad de esta crítica, Francisco Varela (1990) recurre a una potente analogía histórica, señalando que el concepto de información es un "flogisto moderno".

La teoría del flogisto, refutada por Lavoisier, buscaba explicar la combustión postulando la existencia de una sustancia invisible que era liberada al arder. El flogisto nunca existió; fue un marcador de posición para un proceso que se estaba entendiendo al revés: la combustión no es una liberación, sino una combinación con el oxígeno (Kuhn, 2013). Varela utiliza esta historia como una advertencia filosófica precisa. Así como el flogisto era una entidad inventada para explicar un proceso, la "información" es tratada por el cognitivismo como una entidad preexistente y objetiva "ahí fuera" para

explicar el proceso del conocimiento. Se cosifica un accionar dinámico —el enactuar—, convirtiéndolo en una sustancia que puede ser extraída y procesada (Varela, 1990).

La analogía es contundente porque, en ambos casos, la teoría invierte la lógica fundamental del fenómeno. La combustión es una combinación, no una liberación. De modo análogo, el enactivismo propone que el conocimiento no es la captación de información preexistente, sino la creación o enacción de un mundo de significado a través del acoplamiento estructural (Varela, 1990). El significado no se "recoge", emerge en la relación. Mientras la teoría del flogisto colapsó ante el problema empírico del peso, Varela (1990) sostiene que el paradigma informacional colapsa ante el problema filosófico del sentido (*meaning*), pues no puede explicar satisfactoriamente cómo una "información" abstracta puede volverse significativa para un ser vivo sin recurrir a un diseñador externo que inyecte el significado de antemano.

Un defensor del sistema de Floridi podría argumentar, sin embargo, que su teoría evita esta crítica al distinguir entre "información" (semántica) y "datos" pre-semánticos (*dialephora*). Sin embargo, desde la perspectiva enactiva, esto solo desplaza el problema a un nivel más profundo. La pregunta fundamental que el enactivismo le dirige a la ontología de Floridi es: "¿Para quién existe esa *falta de uniformidad*? ¿Quién o qué realiza la operación de distinguir esa diferencia para que se convierta en un *dato*?".

La ontología de Maturana y Varela (1973) sostiene que el universo no viene pre-empaquetado en "datos". Es solo en el momento del acoplamiento estructural que una perturbación del medio se convierte en un dato significativo para un sistema autopoiético, en un acto de creación de sentido (*sense-making*) basado en la necesidad de conservar la vida. El paradigma de Floridi, al postular la primacía de los datos sobre la vida, comete el error categorial de "poner la carreta delante de los bueyes": donde los bueyes son el proceso de vivir, la dinámica autopoiética que activamente "hace emerger" distinciones relevantes de su entorno; y la carreta es la idea misma de un mundo pre-dado y ya parcelado en "datos" objetivos y discretos.

La insuficiencia de la ética informacional: el punto ciego del cognitivismo

La confrontación entre el paradigma informacional y el enactivo es más que un simple ejercicio teórico, pues sus consecuencias se extienden directamente al fundamento de una ética para las tecnologías inteligentes. La ontología que se adopta determina el horizonte de los problemas que se pueden concebir y, por tanto, el alcance de la responsabilidad que se puede asumir. Si bien el marco ético derivado del paradigma dominante es necesario para abordar los desafíos inmediatos de la IA, este análisis argumentará que es radicalmente insuficiente. Dicha insuficiencia se debe a un punto ciego en su ontología fundacional, una incapacidad para reconocer el fenómeno más profundo que la tecnología está desencadenando.

Cartografía del debate hegemónico: gestión, delegación y riesgo existencial. Aunque a primera vista diverso, el debate ético predominante sobre la IA puede articularse en torno a tres grandes frentes, representados por las influyentes figuras de Luciano Floridi, Adela Cortina y Nick Bostrom.[4] A pesar de sus notorias diferencias, sus propuestas convergen en un mismo punto de partida. Al asumir un desarraigo fundamental entre el agente y el artefacto, enmarcan su relación en una lógica de gestión y competencia. Este marco agonístico, centrado en la gestión externa de un "otro" tecnológico, delata su herencia cognitivista común.[5]

Esta lógica de competencia se manifiesta con distintos matices en cada autor. La primera corriente —y nuestro caso paradigmático— es la ética

[4] Es importante reconocer la existencia de otras corrientes de gran relevancia, particularmente aquellas centradas en la justicia algorítmica y la denuncia de sesgos sistémicos (p. ej., Crawford, 2021; Benjamin, 2019; Buolamwini & Gebru, 2018) o los enfoques basados en la ética de las virtudes (p. ej., Vallor, 2016). Sin embargo, y a pesar de su indispensable labor de denuncia de los daños concretos que la IA genera en el presente, estos enfoques no escapan a la lógica de la alteridad aquí criticada.

[5] Este marco teórico encuentra un detallado correlato en el plano geopolítico. La jurista Anu Bradford (2023) documenta la actual «batalla global por regular la tecnología» como una contienda entre tres «imperios digitales» (EE.UU., China y la UE) por el control del artefacto. Su análisis ilustra a gran escala cómo el debate hegemónico se enfoca en esta lógica de gestión externa, omitiendo así la cuestión de la transformación recíproca y la responsabilidad que de ella emana, punto central de la crítica aquí desarrollada.

como gestión de la infoesfera, cuyo máximo exponente es Luciano Floridi. Su propuesta consiste en un marco sistemático de cinco principios: Beneficencia, No maleficencia, Autonomía, Justicia y Explicabilidad. Los cuatro primeros son tomados de la bioética, mientras que la explicabilidad es propuesta específicamente para la IA. Esto se orienta a la gobernanza de los datos y los algoritmos (Floridi, 2024). El foco está puesto en asegurar que los procesos informacionales sean justos, transparentes y seguros, tratando la ética como un problema de diseño y regulación de un nuevo entorno sobre el cual se compite por el control.

La segunda corriente es la ética como límite a la delegación, defendida por Adela Cortina. Desde una perspectiva humanista, su preocupación central no es precisamente la gestión del dato, sino la del artefacto, en vista a la protección de la autonomía y la dignidad humana. El problema ético crucial reside en la diferencia entre hacer uso de los sistemas inteligentes a la hora de tomar decisiones y delegar en esos sistemas inteligentes decisiones significativas para la vida de las personas y de la naturaleza (Cortina, 2024). La tarea ética aquí es trazar una línea infranqueable para asegurar que la agencia moral permanezca en manos humanas, una contienda explícita para preservar un territorio de deliberación frente a un competidor artificial.

La tercera e influyente corriente es la ética como mitigación del riesgo existencial, asociada a Nick Bostrom. Su análisis se proyecta hacia el futuro, centrándose en los peligros de una hipotética Superinteligencia Artificial (AGI). El desafío ético fundamental, desde esta óptica, es el "problema de la alineación de valores": cómo asegurar que una futura inteligencia radicalmente superior a la nuestra adopte objetivos que no sean perjudiciales para la humanidad (Bostrom, 2014). Aquí, el marco agonístico se intensifica hasta volverse existencial: la relación se plantea como una competencia por la supervivencia, donde se debe intentar alinear los objetivos de un rival diametralmente superior para evitar la catástrofe.

Es precisamente este contexto agonístico la estructura conceptual que genera la insuficiencia de la ética informacional. Al enfocar toda la preocupación ética en la gestión de un competidor externo, el debate

hegemónico se vuelve constitutivamente ciego al fenómeno más profundo: la transformación que ocurre dentro de nosotros. La preocupación por la amenaza del "otro" tecnológico impide reconocer que el cambio más radical no es lo que la IA nos *hace*, sino aquello en lo que nos *convertimos* a través de nuestro acoplamiento con ella.

El punto ciego compartido y la necesidad de un nuevo fundamento. Esta insuficiencia no es accidental, sino el síntoma de una aporía en el corazón del paradigma cognitivista. Al llevar el funcionalismo a su conclusión lógica, la tesis floridiana del "divorcio" entre agencia e inteligencia se ve obligada a realizar una compleja operación retórica y ontológica para sostenerse. Para legitimar la existencia de una agencia artificial autónoma pero carente de mente, Floridi (2024) recurre a lo que Perelman y Olbrechts-Tyteca (1989) denominan la técnica de la disociación de las nociones. Mediante esta maniobra, el concepto unitario de "Inteligencia Artificial" es escindido en dos términos jerarquizados. Por un lado, la inteligencia biológica o consciencia es relegada a la categoría de atributo accesorio (Término I), presentándose como un requisito innecesario para la resolución exitosa de tareas. Por otro lado, la agencia funcional —la capacidad de lograr objetivos mediante el procesamiento de datos— es elevada al estatus de "realidad" operativa (Término II). De este modo, la propuesta de Floridi encierra una aporía interna: para que el "divorcio" sea sostenible, debe presuponer una "verdadera inteligencia" cualitativamente distinta de la agencia artificial. Sin embargo, su ontología (Realismo Estructural Informacional), al reducir toda realidad a estructuras de datos, carece de las herramientas para fundamentar dicha diferencia cualitativa, dependiendo así de una noción de inteligencia que su propio sistema no puede sustentar.

Esta reestructuración de la realidad permite a Floridi fundamentar lo que él denomina una "ética blanda" (*soft ethics*). Concebida como un complemento post-cumplimiento (*post-compliance*) a la legislación vigente o "ética dura" (*hard ethics*), la ética blanda se presenta como un marco normativo flexible diseñado para gobernar aquello que la ley no alcanza o no puede prever en la gestión de riesgos inmediatos dentro de la infoesfera (Floridi, 2024). Sin embargo, al instituir una ética para "agentes" que, por definición,

carecen de la condición biológica de posibilidad para la responsabilidad moral —el *sense-making* o creación de sentido—, el sistema genera una paradoja que se hace patente en las propias palabras del autor: "las máquinas [IAs generativas] actuales tienen la inteligencia de una tostadora" (Floridi, 2024, pp. 82-83). A partir de aquí cabe preguntarse con justicia: ¿cómo se explica entonces un cambio histórico tan profundo, que ha movilizado leyes, comisiones internacionales y transformaciones sociales masivas, a partir de tecnologías ontológicamente tan "rudimentarias" como una tostadora?

La dificultad de la ética informacional para responder a esta pregunta y tematizar el fenómeno de fondo se origina en la persistencia de su metáfora fundacional: la distinción "hardware/software". La abstracción funcionalista que subyace a la "ética blanda" tiende a heredar la versión más rígida de esta analogía, tratando al agente humano y al artefacto como entidades conceptualmente distintas y desligadas. Es esta separación la que da lugar a lo que podría llamarse la "coartada del hardware fijo" (Oliver, 1990): una tendencia a asumir que la naturaleza humana permanece estática mientras gestiona herramientas externas. Esta perspectiva ciega al sistema frente al riesgo más profundo: lo que Floridi (2024) celebra como "envoltura" (*enveloping*) —la adaptación del entorno para que la IA funcione con éxito— es, desde la perspectiva enactiva, el inicio de una deriva ontogénica inadvertida. Al adaptar el mundo a agentes funcionales pero estúpidos, corremos el riesgo de simplificar nuestra propia cognición y externalizar funciones de juicio para hacernos legibles ante la máquina, alterando nuestra historia de cambios estructurales recíprocos en la que nos modificamos junto a las tecnologías que creamos y usamos.

La insuficiencia de la ética informacional, en todas sus vertientes, es una consecuencia directa de este punto ciego. Una ética no puede hacerse cargo de un fenómeno que su ontología subyacente no le permite ver. Al estar fundamentada en la gestión de la "carreta" (los datos, las decisiones delegadas, los riesgos futuros), es incapaz de asumir la responsabilidad por la deriva de los "bueyes" (el proceso vivo, nuestra propia transformación ontogénica). Si bien la gestión de los artefactos es crucial, es insuficiente porque el desafío ético más profundo y sin precedentes de la IA no es cómo regulamos el

artefacto, sino cómo navegamos conscientemente la transformación acelerada de nuestro propio modo de ser al acoplarnos con él. Al hacernos conscientes de nuestra plasticidad, emerge una responsabilidad ineludible que el cognitivismo no puede reconocer, y que demanda una nueva fundamentación.

Hacia un principio de responsabilidad ontogénica: fundamentos para una ética enactiva

Si, como se ha venido argumentado, el desafío ético más profundo no reside en la gestión del artefacto, sino en la deriva de nuestro propio ser al acoplarnos con él, entonces se requiere un principio que se haga cargo de esta dinámica co-evolutiva. Esta premisa no es una mera especulación teórica, sino que encuentra un robusto respaldo empírico en la literatura reciente sobre arqueología cognitiva. Como demuestra Malafouris (2025) al analizar la "arqueología del yo", los seres humanos no somos entidades fijas que simplemente usan herramientas, sino procesos de devenir situados y transaccionales. Desde esta óptica, la identidad personal no es un "interior" inmutable, sino un proceso de "in/dividuación enactiva" (*enactive in/dividuation*) que ocurre en el espacio suprapersonal de la interacción material. Esto implica que nuestra subjetividad nunca es puramente individual ni fija, sino que se constituye desde una "perspectiva de la persona situada" (*situated person perspective*), donde los límites del yo se reconfiguran constantemente a través del compromiso material con nuestro entorno técnico (Malafouris, 2025).

El enfoque enactivo, anclado en la biología del conocimiento y reforzado por la evidencia de nuestra constitución protésica, proporciona las bases para articular un nuevo fundamento ético. Entender la cognición como fundamentalmente afectiva —un constante proceso de valoración del mundo en relación con nuestra propia persistencia— implica que la responsabilidad por la forma en que nuestro mundo enactuado evoluciona es una consecuencia directa de nuestra propia naturaleza plástica y permeable. Con esto en mente, puede proponerse el esbozo de un Principio de Responsabilidad Ontogénica: aquel principio en virtud del cual se reconoce que diseñar tecnología es diseñar el entorno al que nos acoplaremos

estructuralmente y, por consiguiente, es participar activamente en el devenir de nuestros futuros modos de ser. Si, como sugiere Malafouris (2025), somos seres "auto-limitados" (*self-bound*) por nuestras prácticas materiales, la responsabilidad ética no se agota en las consecuencias inmediatas de la acción de un artefacto, sino que se extiende a los efectos a largo plazo que el uso de dicho artefacto tiene sobre la deriva ontogénica de los individuos y las culturas.

Este principio, por lo tanto, no busca reemplazar los debates existentes sobre el sesgo o la transparencia, sino complementarlos desde una perspectiva más fundamental. Nos compele a preguntar no solo si un sistema algorítmico tiene sesgos, sino cómo la exposición constante a él moldea nuestra capacidad de comprensión y descubrimiento. La pregunta no es únicamente si un modelo de lenguaje es veraz, sino cómo su uso masivo transforma nuestras habilidades críticas, prácticas democráticas y memoria colectiva, en una deriva que, de ser ciega, corre el riesgo de empobrecer nuestra propia naturaleza.

La responsabilidad ontogénica, si bien novedosa en su fundamentación biológica y arqueológica, dialoga directamente con la tradición de la ética de la tecnología del siglo XX. Su precursor más importante es el "Principio de Responsabilidad" de Hans Jonas (1995). Éste alertó que el poder sin precedentes de la tecnología moderna exigía una nueva ética centrada en nuestra responsabilidad por la supervivencia a largo plazo de la humanidad. Sin embargo, mientras la preocupación de Jonas se centraba en la *posibilidad* de la existencia futura, el principio aquí propuesto se enfoca en la *naturaleza* de esa existencia. La pregunta se desplaza desde *si* habrá humanos en el futuro, hacia *qué tipo de seres humanos* estamos deviniendo a través de nuestro acoplamiento con la tecnología.

De este modo, el principio también recoge las intuiciones de la fenomenología de la tecnología —desarrollada por pensadores como Heidegger (1994), Ihde (1990) y Verbeek (2005)—, que ha descrito extensamente cómo las herramientas no son instrumentos neutros, sino que moldean activamente nuestra experiencia y nuestro modo de ser-en-el-

mundo. El valor agregado del enfoque enactivo es que le otorga a esta descripción fenomenológica un mecanismo biológico concreto: el acoplamiento estructural y la deriva ontogénica son los procesos que explican *cómo* ocurre esta co-constitución entre el ser humano y la técnica.

Este Principio de Responsabilidad Ontogénica, por lo tanto, implica un giro fundamental en el foco de la ética de la IA. La preocupación central se desplaza del artefacto (la IA y sus datos) al acoplamiento (la relación humano-IA que nos transforma). La pregunta ética primordial deja de ser "¿Es justa la IA?" para convertirse en "¿Cómo nos transforma la interacción con la IA?". Este principio no ofrece un conjunto cerrado de reglas, sino que nos compele a adoptar una nueva conciencia sobre nuestro poder como co-creadores de nuestro propio modo de ser, sentando las bases para una ética que esté a la altura del poder transformador de nuestras tecnologías.

Conclusiones

El recorrido de este trabajo comenzó con una analogía astronómica: un giro copernicano para nuestra comprensión de la información y la inteligencia. Hemos cartografiado el ingenioso pero incompleto universo "ptolemaico" del paradigma cognitivista, cuyo centro inmóvil es la información y cuya última expresión se encuentra en el sistemático Realismo Estructural Informacional de Luciano Floridi. A continuación, a través de la mirada enactiva, hemos mostrado que su pilar conceptual —la idea de una información preexistente— representa un "flogisto moderno", un error categorial que invierte la relación fundamental entre la vida y el conocimiento. Frente a este modelo, hemos presentado la alternativa "copernicana" del enactivismo, que desplaza el centro ontológico desde la información hacia el agente vivo y autónomo, argumentando que el conocimiento es la enacción de un mundo de significado en una danza de acoplamiento estructural.

Esta confrontación de paradigmas nos permite reinterpretar la naturaleza misma del éxito tecnológico, conduciéndonos a una conclusión paradójica y profundamente reveladora. El amplio éxito del modelo cognitivista en el campo tecnológico, más allá de validar aquella ontología informacional, puede ser interpretado como una trágica y contundente

muestra de nuestra plasticidad estructural. Que hayamos podido tomar un modelo radicalmente incompleto de nosotros mismos —la mente como un procesador de información—, construir un mundo tecnológico a su imagen y semejanza, y luego transformarnos exitosamente para acoplarnos a él, representa la concreción de una deriva ontogénica insospechada. Muestra que nuestra capacidad de acoplamiento es tan profunda que podemos incluso dar forma a nuestro propio ser para que encaje en un mundo diseñado a partir de una caricatura de nuestra naturaleza.

Esta toma de conciencia, que puede denominarse la "prueba trágica" de nuestra plasticidad, cuestiona firmemente el determinismo tecnológico y nos sitúa frente a una responsabilidad ineludible. Si somos los arquitectos del mundo que nos co-constituye, poseemos también la agencia para transformarlo. De este reconocimiento emerge la necesidad del Principio de Responsabilidad Ontogénica, cuyo valor añadido respecto a las propuestas éticas existentes radica en su capacidad para visibilizar riesgos existenciales que la mera gestión normativa ignora.

Para ilustrar esta diferencia fundamental, consideremos la transformación actual de la práctica educativa frente a la irrupción de los Grandes Modelos de Lenguaje (LLM). La ética informacional se pregunta legítimamente si estas herramientas facilitan el plagio o si sus respuestas son veraces (gestión del artefacto). Sin embargo, la Responsabilidad Ontogénica nos obliga a mirar el cambio estructural en los objetivos mismos del aprendizaje. Al observar cómo las mallas curriculares transitan desde los contenidos hacia las "competencias" —entendidas instrumentalmente como la movilización de recursos para la resolución eficaz de problemas—, vemos cómo la educación se adapta para acoplarse a un mundo donde la generación de texto (ensayos, informes) ya está resuelta tecnológicamente. Si bajo la visión informacional la inteligencia se reduce a la resolución de problemas, y estos problemas ya no requieren de nuestra intervención directa, ¿qué lugar queda para la consciencia que se desarrollaba precisamente al lidiar con ellos? El riesgo no es que la disciplina se vuelva obsoleta, sino que, al dejar de ejercitar los procesos cognitivos que constituían nuestra autonomía intelectual, nos convirtamos en operadores pasivos de un sistema que piensa

por nosotros, reducidos a meros gestores de *outputs* tecnológicos. Lo que este principio aporta, en este caso, es la alerta de que al potenciar nuestros fines mediante la tecnología, corremos el riesgo de modificarlos hasta volverlos irreconocibles. El problema de fondo es que ese camino, esa deriva ontogénica, es lo que nos constituye. Sin asumir esta responsabilidad, podríamos convertirnos en nuestro propio invento: agentes que gestionan resultados eficientes, pero desprovistos de la inteligencia constitutiva del proceso que les daba sentido.

El horizonte que se abre, por tanto, nos enfrenta a una elección fundamental. Por una ruta, se nos ofrece el futuro de la simulación: un mundo donde, siguiendo una deriva ciega, continuamos adaptándonos a un modelo incompleto, con el riesgo de convertirnos en una versión empobrecida de nosotros mismos para ser legibles por nuestras máquinas. Por la otra, se abre la posibilidad de una co-evolución consciente: un futuro donde, reconociendo nuestra profunda condición autopoiética y la naturaleza enactiva de nuestra vida en el lenguaje, diseñamos tecnologías no para reemplazar o escapar de nuestra condición biológica, sino para enriquecerla, en un camino de mayor responsabilidad sobre la forma en que pensamos, diseñamos y vivimos los mundos que, a su vez, nos piensan y nos diseñan de vuelta.

Referencias

Backus, J. (1978). Can programming be liberated from the von Neumann style?: A functional style and its algebra of programs. *Communications of the ACM, 21*(8), 613–641.

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim Code*. Polity Press.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Bradford, A. (2023). *Digital empires: The global battle to regulate technology*. Oxford University Press.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 77-91*. http://proceedings.mlr.press/v81/buolamwini18a.html

Churchland, P. M. (1999). *Materia y conciencia: Introducción contemporánea a la filosofía de la mente* (M. N. Mizraji, Trad.). Gedisa. (Obra original publicada en 1988).

Cortina, A. (2024). *¿Ética o ideología de la Inteligencia Artificial?*. Paidós.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Descartes, R. (2010). *Meditaciones metafísicas*. Editorial Austral.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, and the case for a relational autonomy. *Phenomenology and the Cognitive Sciences, 4*(4), 429–452.

Floridi, L. (2008). A defence of informational structural realism. *Synthese, 161*(2), 219–253. https://doi.org/10.1007/s11229-007-9163-z

Floridi, L. (2010). *Information: A very short introduction*. Oxford University Press.

Floridi, L. (2011). *The philosophy of information*. Oxford University Press.

Floridi, L. (Ed.). (2015). *The onlife manifesto: Being human in a hyperconnected era*. Springer.

Floridi, L. (2024). *Ética de la inteligencia artificial*. Herder.

Floridi, L., & Sanders, J. W. (2004). The method of abstraction. *Minds and Machines, 14*(3), 303–357. https://doi.org/10.1023/B:MIND.0000035461.22188.ee

Fodor, J. A. (1981). *Representations: Philosophical essays on the foundations of cognitive science*. The MIT Press.

Heidegger, M. (1994). La pregunta por la técnica. En *Conferencias y artículos*. Ediciones del Serbal.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent computational abilities. *Proceedings of the National Academy of Sciences, 79*(8), 2554–2558. https://doi.org/10.1073/pnas.79.8.2554

Hurley, S. (1998). *Consciousness in action*. Harvard University Press.

Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Indiana University Press.

Jonas, H. (1995). *El principio de responsabilidad: Ensayo de una ética para la civilización tecnológica* (J. M. Ripalda, Trad.). Herder. (Obra original publicada en 1979).

Kim, J. (2011). *Philosophy of mind* (3rd ed.). Westview Press.

Korbak, T. (2021). Computational enactivism under the free energy principle. *Synthese, 198*(Suppl 1), 2743–2763. https://doi.org/10.1007/s11229-019-02372-w

Kuhn, T. S. (1993). *La revolución copernicana*. Planeta-Agostini. (Obra original publicada en 1957).

Kuhn, T. S. (2013). *La estructura de las revoluciones científicas* (4ª ed.). Fondo de Cultura Económica. (Obra original publicada en 1962).

Machado, A. (1912). *Campos de Castilla*. Renacimiento.

Malafouris, L. (2025). People are STRANGE: Towards a philosophical archaeology of self. *Phenomenology and the Cognitive Sciences, 24*, 685–711. https://doi.org/10.1007/s11097-024-10030-x

Maturana, H. R. (1970). *Biology of cognition* (BCL Report No. 9.0). Biological Computer Laboratory, University of Illinois, Urbana.

Maturana, H. R., & Guilloff, G. (2006). En búsqueda de la inteligencia de la inteligencia. *Psykhe (Santiago), 15*(1), 3–15. https://dx.doi.org/10.4067/S0718-22282006000100001

Maturana, H. R., & Varela, F. J. (1973). *De máquinas y seres vivos: Una teoría sobre la organización biológica*. Editorial Universitaria.

Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding*. Shambhala Publications.

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955, 31 de agosto). *A proposal for the Dartmouth summer research project on artificial intelligence*. Dartmouth College.

Merleau-Ponty, M. (1993). *Fenomenología de la percepción* (J. Cabanes, Trad.). Planeta-Agostini. (Obra original publicada en 1945).

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM, 19*(3), 113–126. https://doi.org/10.1145/360018.360022

Oliver, M. (1990). *The politics of disablement*. Macmillan.

Perelman, C., & Olbrechts-Tyteca, L. (1989). *Tratado de la argumentación: La nueva retórica* (J. Sevilla Muñoz, Trad.). Editorial Gredos.

Rorty, R. (1979). *Philosophy and the mirror of nature*. Princeton University Press.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations*. MIT Press.

Skidelsky, L. (Ed.). (2023). *Introducción a la filosofía de las ciencias cognitivas*. Ediciones Uniandes.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

Turing, A. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.

Varela, F. J. (1990). *Conocer: Las ciencias cognitivas: tendencias y perspectivas*. Gedisa.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. The MIT Press.

Verbeek, P. P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. Penn State University Press.

Villalobos, M., & Palacios, S. (2021). Autopoietic theory, enactivism, and their incommensurable marks of the cognitive. *Synthese, 198*(Suppl 1), S71–S87. https://doi.org/10.1007/s11229-019-02420-5

Villalobos, M., & Silverman, D. (2018). Extended functionalism, radical enactivism, and the autopoietic theory of cognition: Prospects for a full revolution in cognitive science. *Phenomenology and the Cognitive Sciences, 17*(4), 719–739. https://doi.org/10.1007/s11097-018-9573-z

Wheeler, J. A. (1990). Information, physics, quantum: The search for links. En W. H. Zurek (Ed.), *Complexity, entropy, and the physics of information* (pp. 3–28). Addison-Wesley.