

Del conexionismo al deep learning. Discusiones para la validación de teorías cognitivas

From Connectionism to Deep Learning: Discussions on the Validation of Cognitive Theories

Jesús Goenaga Peña¹

jesus.goenagap@autonoma.edu.co

<https://orcid.org/0000-0003-1936-1642>

Luis Fernando Castillo Ossa²

luis.castillo@ucaldas.edu.co

<https://orcid.org/0000-0002-2878-8229>

<https://doi.org/10.22209/rhs.v14n2a03>

Recibido: septiembre 11 de 2025.

Aceptado: mayo 20 de 2026.

Para citar: Goenaga Peña, J., & Castillo Ossa, L. F. (2026). Del Conexionismo al Deep Learning.
Discusiones para la validación de teorías cognitivas. *RHS-Revista Humanismo y Sociedad*, 14(2), 1-16.

<https://doi.org/10.22209/rhs.v14n2a03>

1 Universidad Autónoma de Manizales, Colombia. <https://ror.org/00jfare13>

2 Universidad de Caldas, Colombia. <https://ror.org/049n68p64>

Resumen

Partiendo de la revisión crítica y el análisis del campo de la inteligencia artificial, este artículo busca sostener la tesis de que las teorías cognitivas conexionistas pueden validarse mediante el diseño de modelos cognitivos artificiales basados en técnicas actuales de aprendizaje profundo. Se postula que, si estos modelos se diseñan con base en los principios del paradigma conexionista de las ciencias cognitivas, pueden ofrecer una representación confiable y empíricamente sostenible de la cognición. Se argumenta que la tecnología computacional actual es lo suficientemente sofisticada para emular las condiciones inherentes al enfoque conexionista y a la neurofisiología subyacente a los procesos cognitivos perceptuales. Como ilustración, se examina el proceso de percepción de profundidad visual, defendiendo la idea de que estos modelos autónomos, que incorporan mecanismos de aprendizaje y adaptación, pueden desempeñarse como herramientas experimentales de corroboración o refutación de la teoría cognitiva. A partir de las evidencias encontradas en la literatura, se concluye que el desarrollo de estos modelos es apropiado para la captura de regularidades funcionales de la cognición y constituye una plataforma experimental robusta para la validación de postulados conexionistas.

Palabras clave: conexionismo; sistemas cognitivos artificiales; inteligencia artificial; visión por computador; confirmación científica.

Abstract

Based on a critical review and analysis of the field of artificial intelligence, this article argues that connectionist theories of cognition can be validated through the design of artificial cognitive models using contemporary deep learning techniques. It proposes that models developed in accordance with the principles of the connectionist paradigm in cognitive science can provide a reliable and empirically grounded representation of cognition. The article further contends that current computational technology is sufficiently advanced to emulate the conditions inherent to the connectionist approach and the neurophysiological foundations of perceptual cognitive processes. As an illustrative case, the study examines visual depth perception, arguing that these autonomous models incorporating learning and adaptation mechanisms can function as experimental tools for supporting or challenging cognitive theory. Based on evidence reviewed in the literature, the article concludes that the development of such models is well suited to capturing the functional regularities of cognition and constitutes a robust experimental platform for evaluating connectionist postulates.

Palabras clave: Connectionism, artificial cognitive systems, artificial intelligence, computer vision, scientific confirmation.



Introducción

El presente artículo es producto de un análisis teórico-crítico desarrollado en el marco de la investigación doctoral titulada *Modelo cognitivo artificial para la replicación de la actividad neurofisiológica de la percepción de profundidad*, adscrita a la línea de investigación en Sistemas Cognitivos Artificiales del doctorado en Ciencias Cognitivas de la Universidad Autónoma de Manizales. A partir de la revisión y el análisis del estado actual del conocimiento en Ciencias Cognitivas y Ciencias Computacionales, se plantea como objetivo sostener la tesis de que la teoría conexionista puede obtener evidencia confirmatoria confiable mediante el diseño de modelos cognitivos artificiales basados en las técnicas de aprendizaje profundo actuales.

En este sentido, la búsqueda de una representación adecuada de la cognición ha constituido uno de los principales desafíos en la consolidación formal de las ciencias cognitivas como alternativa explicativa de los fenómenos mentales (Fernández Trespalcios, 1988). La propuesta inicial de los teóricos de la inteligencia artificial (IA) en este campo se orientó al desarrollo de sistemas capaces de imitar la inteligencia humana, en un momento en que el enfoque cognitivo predominante era el simbolismo, centrado en el procesamiento de información mediante símbolos y reglas lógicas (Sloman, 1978).

No obstante, a mediados de la década de 1980, como lo presentan poco después Pitarque y Algarabel (1991), comenzó a adquirir relevancia el enfoque conexionista, sustentado en la idea de que la cognición humana se operacionaliza como un procesamiento computacional distribuido y paralelo. De ahí que, este nuevo enfoque tenía el potencial de capturar las regularidades funcionales de la conectividad neuronal y el funcionamiento cognitivo, a través de modelos computacionales basados en redes neuronales artificiales. Este cambio de paradigma en la ciencia cognitiva dio lugar al desarrollo de sistemas más sofisticados, con capacidad de aprendizaje y adaptación a nuevos entornos y situaciones.

Poco después, las teorías conexionistas perdieron impulso debido, principalmente, a limitantes en la capacidad de procesamiento y del acceso a datos con que se contaba a finales del siglo xx. Este aspecto fue corregido con el desarrollo de la ciencia computacional durante lo que va corrido del siglo xxi, de modo que el diseño de modelos computacionales basados en arquitecturas de redes neuronales y formulados para producir resultados asociados a procesos cognitivos se ha convertido en una carrera ingenieril fértil y de mucha importancia; sin embargo, muchos de estos modelos no se construyen emulando las condiciones naturales de los procesos cognitivos ni su correlato neurofisiológico asociado.

Este trabajo sostiene que diseñar modelos cognitivos siguiendo parámetros bioinspirados y ajustados a la evidencia sobre el funcionamiento natural de la cognición permite probar de manera realista las teorías de la ciencia cognitiva con las que explica su objeto de estudio. Con este propósito, se desarrollan cuatro ejes de análisis sustentados en premisas con respaldo

científico: el primero argumenta que los modelos conexionistas constituyen una representación plausible de la cognición; el segundo muestra cómo los modelos computacionales contemporáneos pueden representar adecuadamente las condiciones del paradigma conexionista; el tercero aborda la neurofisiología de la percepción de profundidad y el modo en que las tecnologías actuales permiten modelar este proceso y, finalmente, el cuarto se centra en el diseño de modelos cognitivos artificiales autónomos como prueba experimental de la teoría conexionista.

Modelos conexionistas como una representación plausible de la cognición

Desde sus inicios, los científicos cognitivos propusieron que la cognición puede entenderse como un proceso de cómputo de información, que implica transformar datos sensoriales en representaciones internas y generar respuestas apropiadas con base en dichas representaciones (Clark, 2001; Pylyshyn, 1980; Sloman, 1978). En este sentido, una característica común a las distintas teorías cognitivas clásicas es que la actividad de conocer —raíz etimológica de cognición: *cognitio* y *cognoscere*— se concibe como procesamiento de información.

Paralelamente, aunque con cierta independencia histórica, la neurociencia ha demostrado que el cerebro humano procesa información relacionada con las funciones cognitivas de manera secuencial y paralela, a través de redes neuronales complejamente interconectadas (Cooper & Shallice, 2010). Hacia finales del siglo xx, la convergencia entre las teorías cognitivas clásicas y los hallazgos de la neurociencia, junto con el desarrollo de la informática, propició el surgimiento del paradigma conexionista en las ciencias cognitivas (Cooper & Shallice, 2010).

El conexionismo, en tanto teoría computacional de la cognición, postula que la mente humana opera mediante la activación y propagación de señales en una red de unidades computacionales, análoga al funcionamiento neuronal (Fernández Trespalacios, 1988). Esto implica un procesamiento de información altamente distribuido y paralelo, semejante al modo en que interactúan las neuronas en el cerebro (Rumelhart & McClelland, 1986).

Desde esta perspectiva, la modelización de procesos cognitivos mediante simulaciones computacionales adquirió un nuevo alcance al establecer como inspiración la conectividad neuronal asociada a la operación cognitiva. Tal premisa se fundamenta en que los modelos conexionistas están diseñados para simular la conectividad y la interacción entre unidades computacionales, interpretadas como análogas a las neuronas (Davis, 1992). Estos modelos buscan representar cómo las conexiones entre unidades cambian con el tiempo a través del aprendizaje, en un proceso semejante al fortalecimiento y la modificación de sinapsis en el cerebro (Corbí & Prades, 1995).

En este sentido, el planteamiento teórico del conexionismo puede y debe representarse computacionalmente, en coherencia con la idea de que estas teorías cognitivas involucran procesamiento de información. Puesto que este enfoque entiende el aprendizaje y la representación como procesos basados en la activación y modificación de conexiones entre nodos o unidades distribuidas en redes, un modelo computacional diseñado bajo estos principios debe reflejar la plasticidad y adaptabilidad del cerebro en la ejecución de tareas cognitivas complejas.

Además, desde su formulación, el enfoque conexionista se ha considerado idóneo para el diseño de modelos artificiales que aprovechan los recursos de la computación científica para desarrollar algoritmos capaces de simular la dinámica de activación y modificación de conexiones neuronales durante el aprendizaje (Fierro, 2011). Este tipo de algoritmos ofrece la oportunidad de verificar la precisión de las teorías cognitivas propuestas desde el conexionismo, puesto que, si el sistema se comporta de manera semejante a la cognición humana, dicho comportamiento constituye evidencia a favor de la teoría.

El uso de estos modelos de computadora se ve justificado por su capacidad de control y precisión en la manipulación de variables y la observación de resultados, dado que constituyen escenarios de experimentación bajo condiciones controladas, lo que permite identificar patrones y reproducir modelos consistentes en distintos momentos y con diferentes conjuntos de datos, aspectos esenciales en la validez y fiabilidad de los hallazgos.

En consecuencia, si un sistema artificial se comporta conforme a las predicciones de la teoría conexionista, ello puede considerarse evidencia en su apoyo, en concordancia con los supuestos filosóficos de que el conocimiento científico se construye mediante un proceso continuo de revisión y actualización, y de que siempre es posible que nuevas evidencias cuestionen o modifiquen teorías existentes.

Adaptabilidad de los modelos computacionales al conexionismo

Los modelos computacionales contemporáneos, como las redes neuronales artificiales (RNA) y los enfoques de procesamiento distribuido y paralelo, han permitido el desarrollo de una línea de investigación de gran relevancia en el campo de la inteligencia artificial y se han consolidado como herramientas fundamentales para representar condiciones funcionales de la cognición (Sarker *et al.*, 2017).

Como se ha señalado, los modelos conexionistas, al aproximarse a la conectividad neuronal y su procesamiento de información, modelan la cognición en coherencia con la evidencia biológica asociada (Yoshimi *et al.*, 2024). Este alcance fue posible gracias al desarrollo computacional de las RNA, modelos compuestos por unidades de procesamiento interconectadas, que reciben señales de entrada, las procesan y generan salidas capaces de simular ciertas características de la actividad neuronal (Lewandowsky & Farrell, 2011). También, con la incorporación de algoritmos de aprendizaje automático, las RNA ajustan sus conexiones para adaptarse a datos específicos, lo que les permite aprender y generalizar patrones, una capacidad esencial en la cognición humana (Farrell & Lewandowsky, 2018).

La relación entre esta representación artificial y la actividad neuronal está respaldada por investigaciones que muestran cómo algunas propiedades de las RNA, como la propagación de información a través de múltiples capas (redes profundas) se asemeja al procesamiento cerebral (Arbib, 2003); o como la plasticidad sináptica, que constituye la capacidad de modificar la fuerza de las conexiones en respuesta al aprendizaje (Donahoe & Dorsel, 1997). De este modo, es procedente sostener que este diseño computacional se ajusta a una representación de la estructura de las redes neuronales biológicas.

Otro aspecto que refuerza la capacidad computacional para adaptarse a los parámetros del conexionismo es la actual disponibilidad de potencia computacional, que permite procesar grandes volúmenes de datos de forma óptima. Este fue un limitante en la emergencia del conexionismo en los 80 porque resulta indispensable para representar la complejidad de la conectividad y del procesamiento cerebral, caracterizado por un manejo masivo de información (Yoshimi *et al.*, 2024). Actualmente, el desarrollo de la ciencia computacional ha permitido avances significativos en capacidad de *hardware* y *software*, permitiendo entrenar modelos de RNA cada vez más grandes y profundos, que gestionan conjuntos de datos de dimensiones que pueden considerarse cercanas a algunas de las que el cerebro humano capta y procesa (Aggarwal, 2018). Esta potencia de cómputo habilita la exploración de modelos conexionistas complejos y realistas que logren representaciones más ajustadas de la actividad cognitiva.

Ahora bien, los modelos conexionistas requieren contar con una arquitectura que replique el carácter distribuido y paralelo del procesamiento cerebral. Estas dos propiedades pueden diferenciarse computacionalmente y permiten un cálculo simultáneo de múltiples entradas de información, lo que incrementa la velocidad y eficiencia del procesamiento (Ayaz *et al.*, 2021).

El carácter distribuido se implementa a través de una red interconectada de nodos, donde cada uno asume de manera autónoma una tarea de procesamiento, a menudo especializada en una porción del problema. Estas unidades comparten datos o resultados parciales, lo que genera transmisiones de información centralizadas o distribuidas según la configuración del sistema.

El paralelismo, por su parte, se refiere a la capacidad sincrónica de procesamiento de múltiples tareas que se ejecutan de manera simultánea en el escenario de un sistema de cómputo (Aloisio *et al.*, 1997). Tanto el procesamiento distribuido como el paralelo cuentan con mecanismos de organización de recursos y coordinación de tareas diversos, puesto que la distribución se apoya en la colaboración de unidades interconectadas, mientras que el paralelismo aprovecha múltiples nodos de procesamiento en un sistema centralizado.

Aunque la ciencia computacional continuará optimizando la potencia y la complejidad de las herramientas disponibles, los modelos basados en RNA, junto con las capacidades de procesamiento distribuido y paralelo, ya han demostrado eficacia para representar la actividad cognitiva bajo parámetros de simulación neurofisiológica.

Modelado computacional de la percepción de profundidad

Para ilustrar el modelado de un proceso cognitivo en la tesis propuesta, resulta pertinente el caso de una función básica dentro de la naturaleza jerárquica de la cognición, como la percepción visual y, específicamente, la percepción de profundidad, un proceso que sirve como fundamento para funciones más complejas como la memoria, el pensamiento y la toma de decisiones.

Asimismo, la percepción es uno de los procesos mejor comprendidos en la psicología y la neurociencia cognitiva, gracias a décadas de investigación empírica (Steinman & Garzia, 2000). De ahí que, iniciar una línea de investigación en torno al desarrollo de modelos computacionales robustos de los procesos perceptuales demarca un primer alcance realista respecto al desarrollo del conocimiento en las disciplinas implicadas; posteriormente, estos modelos pueden incorporarse como componentes en sistemas cognitivos más amplios, que incluyan otros procesos cognitivos más complejos, dado que estos dependen en gran medida de una percepción precisa para funcionar eficazmente.

Ahora bien, como es evidente, las posibilidades del modelado cognitivo de procesos básicos en plataformas artificiales dependen, en buena medida, del desarrollo de la computación científica; por lo que, para modelar la percepción visual, se requiere utilizar las variadas funcionalidades del campo de la visión artificial con fines de investigación básica (Szeliski, 2011). En cambio, iniciar una línea de trabajo en modelos de procesos cognitivos priorizando los mecanismos superiores de la cognición, tales como el pensamiento, el lenguaje o la intencionalidad, sin contar con una base sólida en modelos bioinspirados de la percepción, atención y memoria puede añadir

una complejidad que obstaculice la validación teórica. Por ello, se sugiere un enfoque gradual y estructurado en el desarrollo de modelos computacionales aplicados a la investigación básica, en tanto el objetivo es alcanzar una comprensión profunda de la teoría más allá de soluciones prácticas inmediatas.

La percepción de profundidad visual se define como la capacidad de estimar la relación entre el objeto y el observador visual al percibir el mundo en tres dimensiones (Sweet & Kaiser, 2011). Este proceso perceptual involucra tanto el ajuste del ojo en la captación de la luz, como la interpretación de la distancia entre el objeto y el ojo ocurrente en la corteza visual. El proceso comienza con el ingreso y acomodación de la información lumínica que ingresa al ojo, de modo que la luz que carga con la información de los objetos, tanto cercanos como lejanos, se proyecte con nitidez en la retina. Esta información, asimismo, ingresa de manera binocular, lo que otorga leves diferencias entre las imágenes captadas por ambos ojos y que se emplea para calcular la disparidad retiniana, un indicador crucial de la distancia. La información termina siendo procesada en la corteza visual, donde se interpreta y construye una experiencia perceptual unificada (Kolars, 1972).

Este proceso especializado de la visión —que permite interpretar el entorno con profundidad— implica varios ajustes y transformaciones en el correlato neurofisiológico. En primer lugar, los músculos ciliares modifican la curvatura del cristalino para asegurar que la imagen se proyecte con nitidez en la retina y la distancia entre los objetos afecta directamente la magnitud de este ajuste, lo que da lugar al fenómeno de la acomodación. Luego, la visión binocular y el uso de recursos monoculares contribuyen a que las dos imágenes, ligeramente diferentes, sean proyectadas en cada retina. La diferencia entre ambas proyecciones, la disparidad retiniana, se procesa cuando los impulsos nerviosos que cargan esta información viajan por el nervio óptico hasta el quiasma óptico, estructura en la que las fibras de la mitad nasal de cada retina cruzan hacia el hemisferio cerebral opuesto, mientras que las fibras de la mitad temporal se dirigen al mismo hemisferio, lo cual le permite al cerebro comparar la información de ambos ojos y calcular la disparidad (Forrester *et al.*, 2002).

El procesamiento continúa en la corteza visual primaria (V1), donde la información se descompone en características básicas como la orientación, el contraste y los bordes. No obstante, la percepción de profundidad se desarrolla principalmente en áreas visuales posteriores, como V2, V3, V4 y V5/MT, donde se refinan los cálculos de la disparidad retiniana y se procesan otras claves de profundidad como el sombreado, el tamaño relativo o la superposición, y finaliza con la definición de un mapa de profundidad del entorno visual (Goodale & Milner, 1992; Leuba & Kraftsik, 1994).

Ahora bien, es válido preguntarse si las tecnologías contemporáneas cuentan con la capacidad técnica para capturar y modelar adecuadamente este recorrido neurofisiológico. Si bien los modelos de visión artificial usualmente priorizan la optimización del rendimiento y el

ahorro computacional, dada la naturaleza diferenciada del procesamiento del sistema nervioso respecto del artificial, una revisión de las técnicas actuales de visión por computador sugiere capacidades suficientes en la emulación de distintos momentos del proceso perceptual, particularmente del funcionamiento del sistema ocular y de la conectividad de la corteza visual asociada a la percepción de profundidad (Howard, 2012).

En visión artificial se emplean cámaras y sensores que capturan información visual de manera relativamente semejante a los ojos humanos. Algunos sistemas igualan los límites del ojo humano en la captación de información, lo que los convierte en herramientas de especial interés para el modelado conexionista; sin embargo, también se desarrollan tecnologías que superan los umbrales de valencia que limitan al ojo humano con el fin de abordar tareas de mayor complejidad. Entre estas se incluyen:

- Cámaras estéreo: emplean dos lentes separados para capturar imágenes desde diferentes puntos de vista, simulando la visión binocular humana.
- Cámaras RGB-D: combinan imágenes en color (RGB) con información de profundidad (D) obtenida mediante sensores de tiempo de vuelo (*time-of-flight*, *ToF*) o luz estructurada, que emiten luz infrarroja y miden su rebote en los objetos, lo que permite obtener una representación tridimensional de la escena.
- Cámaras multiespectrales: capturan imágenes en múltiples bandas del espectro electromagnético, incluyendo el visible y el infrarrojo cercano, útiles para detectar características invisibles al ojo humano, con aplicaciones en agricultura y detección de enfermedades vegetales.
- Cámaras de alta velocidad: registran imágenes a gran velocidad y captan eventos que son demasiado rápidos para ser percibidos por el ojo humano, con aplicaciones en análisis del movimiento, biomecánica y colisiones.
- Cámaras hiperespectrales: obtienen imágenes en numerosas bandas espectrales estrechas, lo que permite analizar composición química y reflectancia de objetos, con aplicaciones en detección de minerales, monitoreo ambiental e inspección alimentaria.
- Cámaras de alta resolución: diseñadas para registrar imágenes con gran detalle, esenciales en microscopía e inspección de calidad en manufactura (Pei *et al.*, 2022).

Estas técnicas evidencian los vínculos entre los mecanismos funcionales de la visión artificial y los del ojo humano, incluidos aquellos específicos de la percepción de profundidad. Por ejemplo, respecto al proceso de acomodación, se han desarrollado algoritmos y técnicas que emulan la capacidad de ajustar el enfoque (*autofocus*) en función de la distancia de los objetos, logrando estimar distancia y profundidad a partir de la información binocular, semejante al proceso de disparidad retiniana o estereoscópica del cerebro humano. También, la conectividad en la corteza visual se ha estudiado a través de modelos de redes neuronales convolucionales (CNN) que demuestran capacidad para simular ciertos aspectos del procesamiento visual, incluida la percepción de profundidad (Howard, 2012).

En consecuencia, la evidencia empírica disponible deja inferir que la percepción de profundidad puede modelarse computacionalmente, representando la operación neurofisiológica asociada con un alto grado de similitud.

Modelos cognitivos artificiales autónomos como prueba experimental

Lo novedoso del enfoque de uso aquí propuesto para los modelos cognitivos artificiales no es el desarrollo de nuevas técnicas o diseños algorítmicos que simulen la actividad neuronal y cognitiva, pues se dispone de variadas técnicas compatibles, como hemos visto. No obstante, la tendencia actual en el desarrollo de modelos artificiales con características cognitivas consiste en entregar salidas funcionalmente semejantes al comportamiento humano, pese a diferir en su procesamiento. Si bien esta tendencia está justificada por su capacidad de aprovechamiento, impide el desarrollo de hipótesis teóricas de la cognición, desde una plataforma experimental que constituye una tradición de prueba científica para consolidar conocimiento de nuestros objetos de estudio.

La búsqueda de evidencia que permita confirmar o refutar teorías científicas constituye un objetivo central de la ciencia, bajo la premisa de que la evidencia empírica es el medio fundamental para evaluar la validez de las teorías (Johansson, 2016). Por tanto, la observación y la experimentación funcionan como mecanismos que permiten corroborar o cuestionar supuestos al contrastar sus predicciones con la realidad observada. En este marco, el falsacionismo se ha planteado como el principio filosófico según el cual una teoría científica debe ser susceptible de refutación mediante observaciones o experimentos empíricos (Popper *et al.*, 1978); por ende, ninguna teoría podría probarse de manera definitiva, únicamente corroborada de forma provisional a partir de la evidencia disponible en un momento dado.

Por estas razones, una teoría que no pueda falsarse o que no formule predicciones empíricas específicas no puede considerarse científica. En última instancia, el propósito de la ciencia es ofrecer explicaciones más precisas y generales de los fenómenos naturales, y la falsación es el mecanismo para lograrlo por medio de la revisión constante y actualización de las teorías con base en nuevas evidencias y observaciones (Festa *et al.*, 2005).

Para el caso del diseño de modelos cognitivos artificiales, si su comportamiento coincide con las predicciones de la teoría cognitiva conexionista, pueden considerarse como prueba empírica y evidencia a su favor (Festa *et al.*, 2005). Pero, este tipo de evidencia no invalida necesariamente teorías alternativas, puesto que el hecho de que el modelo se comporte según lo

esperado no implica que la teoría sea la única explicación posible; otros enfoques también podrían dar cuenta de los resultados observados. Esta forma de validación se conoce como confirmación positiva o verificación, y supone que la teoría se considera apoyada cuando los resultados experimentales u observacionales coinciden con sus predicciones.

Puesto que los modelos conexionistas permiten emular la dinámica de activación y modificación de conexiones neuronales durante el despliegue de información en arquitecturas de red, su comportamiento puede compararse con el del sistema cognitivo humano como apoyo o cuestionamiento de la teoría (Popper, 1963/2014). Sin embargo, como ya se dijo, siempre existe la posibilidad de que nuevas observaciones o experimentos futuros cuestionen o modifiquen la teoría. Aun así, la confirmación positiva es fundamental para el desarrollo del conocimiento científico, puesto que constituye un medio para evaluar y validar teorías.

Ahora bien, en los diseños computacionales existe el riesgo de una alta probabilidad de encontrar la evidencia que se está buscando; es decir, si los modelos se diseñan cuidadosamente para representar una arquitectura que les permita ejecutar tareas o resolver problemas específicos, tienden resultar efectivos gracias a la programación precisa de sus parámetros y al control riguroso de variables. Entonces, resulta razonable esperar un desempeño comparable al del sistema cognitivo humano si se han programado exhaustivamente con ese propósito.

Sin embargo, el modelado de procesos cognitivos desde un enfoque neuromórfico, impone cierta rigidez inherente a los parámetros y limita la libertad de acción del programador, lo que repercute en una garantía para que la salida del modelo sea comparable con la del sistema cognitivo humano. En otras palabras, el modelado cognitivo con fines de validación de teoría debe diseñarse sin emplear arquitecturas que faciliten la ejecución de tareas desde un enfoque de optimización, salvo que estas técnicas respondan a los parámetros neurofisiológicos que se buscan modelar (Mesulam, 1990).

El objetivo de este ejercicio es lograr una simulación lo más cercana posible a la forma en que los órganos sensoriales y el sistema nervioso captan y procesan la información, lo que exige guiar la acción del programador desde una serie de parámetros definidos *a priori*, inspirados en las condiciones de conectividad observadas en el cerebro, así como implementar procesamiento distribuido y paralelo en fases claramente diferenciadas. Aunque estas restricciones dificulten la búsqueda de soluciones computacionalmente más simples, garantizan que el modelo no se limite a generar salidas semejantes a las humanas, sino que reproduzca un procesamiento análogo.

Ahora bien, el diseño de modelos computacionales basados en el correlato neurofisiológico de los procesos cognitivos, programados de principio a fin, ha sido aplicado por diversos enfoques en neurociencia computacional (Mesulam, 1990); empero, su validez suele alcanzar el nivel descriptivo, ya que la programación detallada de un modelo suele llevar a que emule exactamente el comportamiento esperado, siguiendo patrones controlados desde el inicio. En

tal caso, la evidencia computacional consiste en una representación de la actividad neurofisiológica, pero no podría considerarse una confirmación de la teoría subyacente.

Por el contrario, si el modelo se configura para operar de manera autónoma, utilizando parámetros que reproduzcan las condiciones experimentadas por el cerebro, se le permite aprender sin intervención directa y acaba generando resultados semejantes al comportamiento humano y desarrollando arquitecturas de conectividad similares a las redes neuronales biológicas, se demostraría que el sistema ha seguido un patrón computacional análogo al cerebral para producir una salida cognitiva eficiente sin haber sido programado de manera exhaustiva desde el inicio para hacerlo. Este tipo de modelo de aprendizaje autónomo constituiría una confirmación positiva del conexionismo, al concebir la cognición como un procesamiento de cómputo que, bajo parámetros bioinspirados, genera autónomamente una actividad cognitiva.

En este contexto, para que un modelo cognitivo artificial constituya una prueba experimental de la teoría conexionista, debe implementar algoritmos de aprendizaje profundo que le permitan adaptarse y generalizar aprendizajes a partir de los datos de entrada, identificando patrones y relaciones para ejecutar tareas progresivamente más complejas y producir predicciones precisas (Goodfellow *et al.*, 2016). Además, debe contar con una arquitectura de redes neuronales adecuada que procese la información ajustando dinámicamente sus pesos, a partir de parámetros definidos *a priori* de acuerdo con las operaciones subyacentes al procesamiento de información en el cerebro.

Finalmente, lo que confiere a estos modelos la capacidad de constituir un escenario experimental para la contrastación de teorías científicas es la inclusión de mecanismos de parametrización orientados a la autonomía, es decir, a la capacidad del sistema para aprender y adaptarse a partir de los datos sin intervención constante o supervisión directa (Goodfellow *et al.*, 2016).

Más específicamente, la autonomía en modelos artificiales que implementan aprendizaje profundo se ha logrado mediante mecanismos de aprendizaje continuo, sin limitarse al entrenamiento inicial del modelo, sino reforzándolo para que logre adaptarse y optimizar su desempeño frente a nuevos datos. De esta forma, se ajustan los parámetros internos, como los pesos de las conexiones, en función de la retroalimentación obtenida en tareas específicas, mediante algoritmos de optimización orientados a incrementar su precisión y eficiencia (Goodfellow *et al.*, 2016).

Debates actuales y desafíos en la validación conexionista

Es necesario reconocer que el valor confirmatorio del diseño de modelos cognitivos artificiales enfrenta retos importantes en la neurociencia computacional y en la inteligencia artificial contemporánea. Entre estos destacan las críticas sobre correspondencia entre redes profundas y actividad cerebral, puesto que estudios recientes han mostrado que las conclusiones sobre similitud entre modelos de aprendizaje profundo y datos neuronales son frágiles y dependen en gran medida de los conjuntos de datos utilizados, la definición del entrenamiento y las métricas seleccionadas, lo que obliga a relativizar la fuerza de estas comparaciones (Soni *et al.*, 2024). Por ello, algunos autores han planteado la necesidad de formular marcos integradores que incluyan múltiples modalidades para fortalecer la validez ecológica de las comparaciones y hacer más explícitos los límites de los modelos conexionistas (Oota *et al.*, 2024).

Otro reto central reside en la plausibilidad neurofisiológica de los mecanismos de aprendizaje. La distancia entre métodos de entrenamiento ingenieriles, como la retropropagación y las reglas de plasticidad neuronal, es en la actualidad un punto crítico, que ha llevado a explorar algoritmos locales que preservan la plausibilidad biológica sin sacrificar eficacia computacional (Lv *et al.*, 2025).

Asimismo, aunque la computación neuromórfica se ha consolidado como un campo promotor para alcanzar sistemas más dinámicos, asíncronos y de bajo consumo, la eficiencia energética y la capacidad de procesamiento en tiempo real siguen representando desafíos importantes. En este ámbito, los esfuerzos de codiseño entre algoritmos, *software* y *hardware* siguen permitiendo avances significativos orientados a la construcción de sistemas más eficientes y escalables, capaces de acercar las condiciones experimentales de laboratorio a escenarios de procesamiento continuo y en entornos reales (Yao *et al.*, 2024; Zhou *et al.*, 2024).

Aun con estos desafíos, el creciente desarrollo de la ciencia computacional y su refinamiento en soluciones de algoritmos locales y continuos —implementados en *hardware* neuromórfico— ofrecerá mejores herramientas y paulatinamente más ajustadas a los requerimientos de un modelado computacional de la cognición. Al mismo tiempo, desde las ciencias cognitivas se hace necesario desarrollar protocolos de validación que combinen métricas conductuales, neurofisiológicas y de eficiencia energética en entornos de validez ecológica para que los modelos cognitivos artificiales puedan consolidarse como plataformas robustas para la confirmación de teoría cognitiva.

Conclusiones

El conexionismo sitúa la cognición humana como el resultado de redes de unidades interconectadas que actúan de manera compleja; sumado a esto, ofrece un marco teórico sólido para comprender los procesos cognitivos y orientar el diseño de modelos artificiales capaces de reproducir, al menos en parte, la dinámica de los procesos cognitivos con plausibilidad neurofisiológica. A diferencia del inicial enfoque simbolista de las ciencias cognitivas, el conexionismo plantea que la cognición se representa mejor como una aproximación computacional de la conectividad neuronal natural, lo que otorga a sus modelos un carácter empíricamente defendible.

Los avances en potencia de cómputo y en algoritmos de aprendizaje profundo han abierto un terreno fértil para que los modelos cognitivos artificiales se consoliden como instrumentos de prueba para confirmar teorías cognitivas conexionistas. El caso de la percepción de profundidad muestra que es posible emular con suficiente fidelidad los procesos específicos asociados, lo que refuerza la idea de que estos modelos no solo reproducen resultados funcionales, sino que pueden aproximarse al modo en que el cerebro organiza y procesa la información.

Más allá de la simulación, la autonomía inherente de estos sistemas constituye quizá su mayor valor científico. La capacidad de aprender, adaptarse y evolucionar sin intervención humana constante los convierte en escenarios experimentales idóneos para poner a prueba las premisas conexionistas. En este orden de ideas, los modelos cognitivos artificiales no deben entenderse únicamente como herramientas tecnológicas, sino como plataformas epistemológicas que amplían las posibilidades de la investigación en ciencias cognitivas.

Referencias

- **Aggarwal**, C. C. (2018). *Neural networks and deep learning: A textbook*. Springer.
<https://doi.org/10.1007/978-3-319-94463-0>
- **Aloisio**, G., Boichicchio, M. A., La Scala, M., & Sbrizzai, R. (1997). A distributed computing approach for real-time transient stability analysis. *IEEE Transactions on Power Systems*, 12(2), 981–987.
<https://doi.org/10.1109/59.589801>
- **Arbib**, M. A. (Ed.). (2003). *The handbook of brain theory and neural networks* (2nd ed.). MIT Press.
- **Ayaz**, H., Asgher, U., & Paletta, L. (Eds.). (2021). *Advances in neuroergonomics and cognitive engineering: Proceedings of the AHFE 2021 Virtual Conferences on Neuroergonomics and Cognitive Engineering, Industrial Cognitive Ergonomics and Engineering Psychology, and Cognitive Computing and Internet of Things, July 25-29, 2021, USA* (Vol. 259). Springer.
<https://doi.org/10.1007/978-3-030-80285-1>

- **Clark, A.** (2001). *Mindware: An introduction to the philosophy of cognitive science*. Oxford University Press.
- **Cooper, R. P. & Shallice, T.** (2010). Cognitive neuroscience: The troubled marriage of cognitive science and neuroscience. *Topics in Cognitive Science*, 2(3), 398–406.
<https://doi.org/10.1111/j.1756-8765.2010.01090.x>
- **Corbí, J. E. & Prades, J. L.** (1995). El conexionismo y su impacto en la filosofía de la mente. En F. Broncano (Ed.), *La mente humana* (pp. 151–174). Trotta.
- **Davis, S.** (Ed.) (1992). *Connectionism: Theory and practice*. Oxford University Press, USA.
- **Donahoe, J. W. & Dorsel, V. P.** (Eds.) (1997). *Neural network models of cognition: Biobehavioral foundations*. Elsevier.
- **Farrell, S. & Lewandowsky, S.** (2018). *Computational modeling of cognition and behavior*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316272503>
- **Fernández Trespalacios, J. L.** (1988). El conexionismo. *Aldaba: Revista del Centro Asociado a la UNED de Melilla*, (11), 25–40. <https://doi.org/10.5944/aldaba.11.1988.20132>
- **Festa, R., Aliseda, A., & Peijnenburg, J.** (Eds.). (2005). *Confirmation, empirical progress, and truth approximation: Essays in debate with Theo Kuipers, Volume 1*. Brill.
<https://doi.org/10.1163/9789401202855>
- **Fierro, M.** (2011). El desarrollo conceptual de la ciencia cognitiva. Parte I. *Revista Colombiana de Psiquiatría*, 40(3), 519–533. [https://doi.org/10.1016/S0034-7450\(14\)60144-X](https://doi.org/10.1016/S0034-7450(14)60144-X)
- **Forrester, J. V., Dick, A. D., McMenamin, P. G., & Roberts, F.** (2002). Anatomy of the eye and orbit. En J. V. Forrester, A. D. Dick, P. G. McMenamin, & F. Roberts, *The eye: Basic sciences in practice* (2nd ed., pp. 66–90). Saunders.
- **Goodale, M. A. & Milner, A. D.** (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25. [https://doi.org/10.1016/0166-2236\(92\)90344-8](https://doi.org/10.1016/0166-2236(92)90344-8)
- **Goodfellow, I., Bengio, Y., & Courville, A.** (2016). *Deep learning*. MIT Press.
- **Howard, I. P.** (2012). *Perceiving in depth, volume 1: Basic mechanisms*. Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199764143.001.0001>
- **Johansson, L.-G.** (2016). *Philosophy of science for scientists*. Springer.
<https://doi.org/10.1007/978-3-319-26551-3>
- **Kolers, P. A.** (1972). *Aspects of motion perception*. Pergamon Press.
- **Leuba, G. & Kraftsik, R.** (1994). Changes in volume, surface estimate, three-dimensional shape and total number of neurons of the human primary visual cortex from midgestation until old age. *Anatomy and Embryology*, 190(4), 351–366. <https://doi.org/10.1007/BF00187963>
- **Lewandowsky, S. & Farrell, S.** (2011). *Computational modeling in cognition: Principles and practice*. SAGE Publications. <https://doi.org/10.4135/9781483349428>
- **Lv, C., Xu, J., Lu, Y., Wang, X., Wang, Z., Xu, Z., Yu, D., Du, X., Zheng, X., & Huang, X.** (2025). Dendritic localized learning: Toward biologically plausible algorithm. In A. Singh, M. Fazel, D. Hsu, S. Lacoste-Julien, F. Berkenkamp, T. Maharaj, K. Wagstaff, & J. Zhu (Eds.), *Proceedings of the 42nd International Conference on Machine Learning* (Vol. 267, pp. 41682–41700). PMLR.
<https://proceedings.mlr.press/v267/lv25c.html>

- **Mesulam, M.-M.** (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology*, 28(5), 597–613.
<https://doi.org/10.1002/ana.410280502>
- **Oota, S. R., Chen, Z., Gupta, M., Surampudi, B. R., Jobard, G., Mounier, A., & Hale, J.** (2024). Deep neural networks and brain alignment: Brain encoding and decoding (survey). *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=n9cAiCGkGv>
- **Pei, Y., Li, Z., Li, B., Zhao, Y., He, H., Yan, L., Li, X., Wang, J., Zhao, Z., Sun, Y., Zhou, Z., Zhao, J., Guo, R., Chen, J., & Yan, X.** (2022). A multifunctional and efficient artificial visual perception nervous system with Sb₂Se₃/CdS-core/shell (SC) nanorod arrays optoelectronic memristor. *Advanced Functional Materials*, 32(29), Article 2203454. <https://doi.org/10.1002/adfm.202203454>
- **Pitarque, A. & Algarabel, S.** (1991). El conexionismo como marco de simulación: Aplicación a una tarea de facilitación semántica. *Cognitiva*, 3(2), 165–186.
- **Popper, K., Adorno, T. W., Dahrendorf, R., & Habermas, J.** (1978). *La lógica de las ciencias sociales*. Grijalbo.
- **Popper, K.** (1963-2014). *Conjectures and refutations: The growth of scientific knowledge*. Routledge.
- **Pylyshyn, Z. W.** (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, 3(1), 111–132. <https://doi.org/10.1017/S0140525X00002053>
- **Rumelhart, D. E. & McClelland, J. L.** (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1: Foundations*. MIT Press.
- **Sarker, M. T., Noor, S., & Acharjee, U. K.** (2017). Basic application and study of artificial neural networks. *SK International Journal of Multidisciplinary Research Hub*, 4(4), 1–12.
- **Szeliski, R.** (2011). *Computer vision: Algorithms and applications*. Springer.
<https://doi.org/10.1007/978-1-84882-935-0>
- **Sloman, A.** (1978). *The computer revolution in philosophy: Philosophy, science, and models of mind*. Harvester Press.
- **Soni, A., Srivastava, S., Kording, K., & Khosla, M.** (2024). Conclusions about neural network to brain alignment are profoundly impacted by the similarity measure. *bioRxiv*.
<https://doi.org/10.1101/2024.08.07.607035>
- **Steinman, S. B. & Garzia, R. P.** (2000). *Foundations of Binocular Vision: A Clinical Perspective*. McGraw-Hill.
- **Sweet, B. & Kaiser, M.** (2011). Depth perception, cueing, and control. *AIAA Modeling and Simulation Technologies Conference*. AIAA. <https://doi.org/10.2514/6.2011-6424>
- **Yao, M., Richter, O., Zhao, G., Qiao, N., Xing, Y., Wang, D., Hu, T., Fang, W., Demirci, T., De Marchi, M., Deng, L., Yan, T., Nielsen, C., Sheik, S., Wu, C., Tian, Y., Xu, B., & Li, G.** (2024). Spike-based dynamic computing with asynchronous sensing-computing neuromorphic chip. *Nature Communications*, 15, Article 4464. <https://doi.org/10.1038/s41467-024-47811-6>
- **Yoshimi, J., Tosi, Z., Hotton, S., Beckmann, P., Cain, E., Gordon, C., Noelle, D. C., & Meyer, T.** (2024). *Neural networks in cognitive science*. https://jeffryoshimi.net/downloads/NeuralNetsCogSci_2023.pdf
- **Zhou, C., Zhang, H., Yu, L., Ye, Y., Zhou, Z., Huang, L., Ma, Z., Fan, X., Zhou, H., & Tian, Y.** (2024). Direct training high-performance deep spiking neural networks: A review of theories and methods. *Frontiers in Neuroscience*, 18. <https://doi.org/10.3389/fnins.2024.1383844>