

Inteligencias Paralelas

Puentes entre las redes neuronales biológicas y las redes neuronales computacionales

DOI: 10.29236/sistemas.n177a7

Resumen

La relación entre las redes neuronales biológicas y las redes neuronales artificiales constituye uno de los diálogos científicos más fértiles de la actualidad. Aunque surgieron en contextos distintos, ambos sistemas buscan comprender y modelar la inteligencia desde perspectivas complementarias: la biología mediante la exploración de sinapsis, plasticidad y circuitos funcionales, y la computación mediante arquitecturas matemáticas capaces de aprender a partir de datos. Este artículo examina los principios fundamentales de cada enfoque y analiza sus puentes conceptuales, evitando reducir uno al otro. Se abordan temas como la organización del conectoma, la dinámica sináptica, la evolución histórica de las redes artificiales y su capacidad para procesar información mediante optimización estadística. Asimismo, se exploran escenarios de convergencia como las interfaces cerebro-computador, la computación neuromórfica y las analogías emergentes entre el olvido biológico y los métodos de *machine unlearning*. A través de este análisis, se muestra que la interacción entre neurociencia e inteligencia artificial no persigue imitación estructural, sino inspiración funcional y expansión mutua. La comparación crítica entre ambos sistemas permite comprender mejor sus alcances, límites y posibilidades, y abre nuevas vías para el desarrollo de tecnologías que complementen, y no sustituyan, las capacidades humanas.

Palabras claves

Neurociencia, Inteligencia Artificial, Redes Neuronales, Plasticidad Sináptica, *Machine Unlearning*

Introducción

“Lo que hoy llamamos inteligencia artificial no es más que el esfuerzo por entender la inteligencia humana.”

— Marvin Minsky, uno de los fundadores de la IA.

La búsqueda por comprender la inteligencia ha conducido, desde mediados del siglo XX, a un diálogo cada vez más estrecho entre neurociencia y ciencias de la computación. Dos trayectorias que nacieron separadas —una enfocada en desentrañar la arquitectura viva del cerebro y otra en construir modelos matemáticos capaces de aprender— han confluído gradualmente en un terreno común. La neurociencia ha revelado que la mente humana es un sistema profundamente interconectado, organizado en redes y subredes cuya dinámica puede describirse mediante principios de conectividad estructural y funcional (Bassett & Sporns, 2017). Al mismo tiempo, la inteligencia artificial moderna se ha expandido desde modelos simples como el perceptrón (McCulloch & Pitts, 1943; Rosenblatt, 1958) hacia arquitecturas profundas capaces de interpretar imágenes, producir lenguaje y resolver tareas complejas en dominios ampliamente variables (LeCun et al., 2015).

En las últimas décadas, logros como la victoria de *AlphaGo* sobre el

campeón mundial de Go, el desempeño sobresaliente de agentes computacionales en videojuegos de estrategia en tiempo real y los avances de los vehículos autónomos han mostrado que las máquinas pueden aprender patrones sutiles sin depender de reglas explícitas. Estos sistemas están construidos sobre redes neuronales profundas entrenadas con grandes volúmenes de datos, utilizando principios matemáticos que, aunque inspirados remotamente en el cerebro, operan bajo lógicas muy distintas (Goodfellow et al., 2016). En lugar de reproducir procesos biológicos, estas redes implementan optimizaciones numéricas que permiten ajustar millones de parámetros para lograr comportamientos adaptativos.

Mientras la informática avanza en su propio marco conceptual, la neurociencia sigue describiendo el cerebro con un nivel de detalle cada vez más fino. El conocimiento actual sobre la sinapsis, los mecanismos de potenciación a largo plazo y depresión sináptica, y las bases moleculares del aprendizaje revelan un sistema plástico, químico y altamente especializado (Kandel et al., 2013). Las investigaciones sobre el conectoma y la organización en redes funcionales demuestran que el procesamiento cognitivo emerge de la interacción coordinada de regiones distantes, en esca-

las temporales que van de los milisegundos a varios segundos (Edlow & Menon, 2024). A diferencia de las redes artificiales, cuya arquitectura es diseñada y definida de antemano, las redes biológicas se reorganizan permanentemente en función de la experiencia y el entorno.

A pesar de sus diferencias fundamentales, ambos campos convergen en conceptos que resuenan de manera profunda. Tanto las redes biológicas como las artificiales operan mediante conexiones entre unidades elementales, y en ambos casos el aprendizaje surge de la modificación de esas conexiones. Pero la semejanza es conceptual, no estructural. Las neuronas biológicas son entidades complejas que integran señales electroquímicas, participan en mecanismos oscilatorios y dependen del soporte de células gliales; las unidades artificiales, en contraste, son funciones matemáticas sencillas que reciben entradas numéricas, aplican transformaciones y generan salidas. Allí donde el cerebro utiliza neurotransmisores, ritmos oscilatorios, neuromodulación y un entorno químico dinámico, las redes artificiales utilizan ecuaciones lineales y funciones de activación como ReLU o sigmoide.

La comparación se vuelve aún más rica cuando se examina el papel del olvido. En neurociencia, el olvido cumple funciones adaptativas esenciales: evita la saturación del

sistema, optimiza la flexibilidad cognitiva y permite reorganizar el conocimiento en función de nuevas experiencias. Procesos como la poda sináptica en el desarrollo o la depresión sináptica en circuitos maduros ilustran cómo el cerebro elimina conexiones para mejorar su eficiencia (Kandel et al., 2013). En un plano completamente distinto, pero conceptualmente cercano, la inteligencia artificial contemporánea ha comenzado a explorar el machine unlearning, un conjunto de técnicas que buscan permitir a los modelos eliminar información previamente aprendida sin reconstruirlos desde cero. Esta aproximación surge de la necesidad de cumplir requisitos de privacidad y de corregir sesgos o dependencias indeseadas en el entrenamiento, y representa un desafío técnico aún abierto (Xu et al., 2023). Aunque la analogía es limitada, resulta llamativo que la computación esté empezando a enfrentar —desde otra lógica— problemas que la biología resolvió hace millones de años.

La intersección entre neurociencia e inteligencia artificial no pretende fusionar ambos dominios, sino permitir que se iluminen mutuamente. La biología aporta modelos de organización complejos, eficientes y sorprendentemente robustos; la computación ofrece abstracciones matemáticas que permiten explorar hipótesis, visualizar patrones y diseñar sistemas adaptativos. El diálogo entre ambos campos abre la puerta a nuevas preguntas sobre el

aprendizaje, la memoria, el olvido y la representación de la información, y fortalece la comprensión de dos sistemas que, aunque profundamente distintos, comparten la aspiración de explicar cómo emerge la inteligencia.

Fundamentos Biológicos de las Redes Neuronales Naturales (RNB)

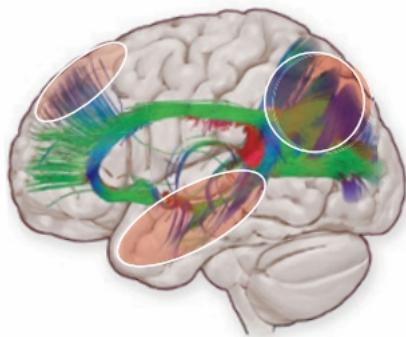
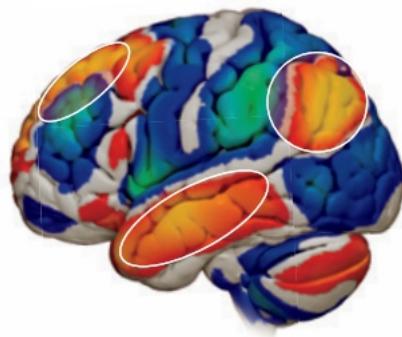
Comprender la organización del cerebro humano es adentrarse en uno de los sistemas más complejos que existen en la naturaleza. Su arquitectura, formada por aproximadamente ochenta y seis mil millones de neuronas y billones de sinapsis, no sigue un diseño estático ni rígido: evoluciona, se reorganiza, se fortalece y se debilita en función de la experiencia. Mientras que las redes artificiales se construyen a partir de ecuaciones y parámetros predefinidos, las redes biológicas emergen de interacciones bioquímicas y electrofisiológicas cuya riqueza desafía cualquier abstracción matemática.

La unidad fundamental de comunicación neuronal es la sinapsis. Lejos de ser un simple punto de contacto entre neuronas, la sinapsis constituye un espacio altamente especializado en el que convergen señales químicas y eléctricas que modelan el flujo de información. Existen sinapsis eléctricas —más rápidas, basadas en uniones gap— y sinapsis químicas —más lentas pero modulables—, además del

modelo sináptico tripartito donde intervienen los astrocitos, añadiendo una capa adicional de regulación (Kandel et al., 2013). La presentación Redes Neuronales Biológicas ilustra cómo la sinapsis es mucho más que una conexión: es un microambiente que sincroniza, amplifica o inhibe información de forma dinámica.

A partir de estas unidades microscópicas emergen estructuras mesoscópicas como las columnas corticales, bloques funcionales que organizan el procesamiento sensorial y cognitivo. Estas columnas no son entidades aisladas, sino nodos altamente interconectados en redes distribuidas. En regiones como la corteza sensorial primaria, por ejemplo, convergen señales del mundo externo que luego se distribuyen jerárquicamente hacia áreas superiores encargadas de integrar y reinterpretar esa información (Mountcastle, 1997). Es en esta integración local y global donde se esconde gran parte de la sofisticación cognitiva humana.

En un nivel aún más amplio, el conectoma cerebral constituye la arquitectura global de conexiones estructurales y funcionales que vinculan regiones distantes del cerebro. Estudios de neuroimagen han permitido observar cómo estas redes presentan patrones robustos de conectividad, organizados en módulos funcionales que interactúan mediante hubs que actúan como centros logísticos de información

B Map of Anatomical Connectivity**C Map of Functional Connectivity**

Fox MD. N Engl J Med. 2018 Dec 6;379(23):2237-2245

(Sporns, 2011). Dichas redes presentan propiedades de eficiencia, resiliencia y organización jerárquica que han inspirado modelos computacionales de procesamiento distribuido (Bassett & Sporns, 2017).

El sistema visual es uno de los ejemplos más elocuentes de organización jerárquica en el cerebro. Señales provenientes de la retina viajan al tálamo y luego a la corteza visual primaria, donde neuronas especializadas responden a líneas, orientaciones o movimientos específicos. A medida que la información asciende hacia áreas superiores, se combinan características simples hasta construir representaciones complejas como rostros, objetos o escenas (Kandel et al., 2013). Esta lógica jerárquica recuerda, en cierta medida, la manera en que las redes convolucionales procesan imágenes, aunque el paralelismo es conceptual y no estructural.

Otro ejemplo es la red de memoria episódica que vincula el hipocampo con la corteza prefrontal. Esta red permite registrar, consolidar y recuperar experiencias personales y, al mismo tiempo, suprime información irrelevante para evitar la saturación. Su funcionamiento muestra que la memoria biológica no es un archivo estático, sino una reconstrucción dinámica y en constante actualización, afectada por emociones, contexto y expectativas.

A gran escala, las redes funcionales identificadas mediante resonancia magnética funcional —como la red por defecto, la red fronto-parietal o la red sensoriomotora— revelan cómo el cerebro coordina regiones distantes para ejecutar tareas complejas como la introspección, la planificación o la atención sostenida (Fox, 2018). Estas redes funcionan como sistemas cooperativos donde la activación de una región tiene repercusiones sistémicas en muchas otras.

A pesar de su sofisticación, el cerebro humano no es un procesador perfecto. Su actividad está atravesada por ruido neuronal, variabilidad intrínseca y limitaciones energéticas. Este ruido, lejos de ser un problema, favorece la adaptabilidad y evita el determinismo rígido, permitiendo que el sistema explore múltiples configuraciones funcionales (Faisal et al., 2008). De hecho, mientras que los grandes modelos de IA consumen cantidades masivas de energía, el cerebro opera con apenas unos 20 vatios, una eficiencia que aún no tiene paralelo artificial.

Su naturaleza electroquímica, aunque genera un inmenso repertorio de posibilidades, introduce también límites: las señales tardan milisegundos en propagarse, la reorganización sináptica requiere tiempo biológico y la plasticidad está influenciada por factores hormonales, emocionales y genéticos. Lejos de ser una desventaja, esta lentitud relativa permite que el cerebro priorice estabilidad, aprendizaje contextual y resiliencia.

Finalmente, el cerebro es producto de millones de años de evolución, no de diseño racional. Su arquitectura refleja una historia de adaptaciones, compromisos y reutilización de estructuras que, aunque eficiente, no sigue los principios de optimización computacional que orientan la ingeniería moderna (Edelman, 1987). Esta diferencia epistemológica es clave para en-

tender por qué compararlo con una red artificial puede ser útil metafóricamente, pero insuficiente en términos estructurales.

Fundamentos Computacionales de las Redes Neuronales Artificiales

La idea de construir modelos computacionales inspirados en el cerebro no es nueva; surgió en el corazón de la cibernetica, cuando matemáticos y neurofisiólogos empezaron a preguntarse si los principios de la actividad neuronal podían formalizarse en ecuaciones. Esta intuición inicial dio origen a un recorrido fascinante. Algunas metáforas sobrevivieron, otras se transformaron o se abandonaron, pero el objetivo persistió: diseñar sistemas capaces de aprender. Las redes neuronales artificiales, tal como se conocen hoy, son el resultado acumulado de décadas de experimentación científica, avances tecnológicos y revoluciones conceptuales.

El primer intento formal de capturar el comportamiento de una neurona en términos matemáticos fue el trabajo de McCulloch y Pitts (1943), quienes describieron una “neurona lógica”: una unidad que integraba entradas binarias y generaba una salida basada en reglas formales.

Aunque extraordinariamente simple, este modelo inauguró la posibilidad de representar procesos cognitivos mediante circuitos computacionales, Ver Figura 1.

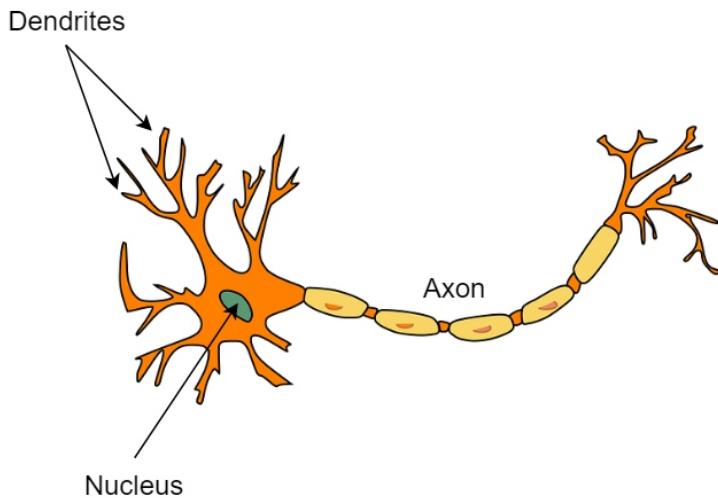


Figura 1. Anatomía de una neurona. Adaptado de Towards Data Science. <https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc>

Años después, Rosenblatt (1958) propuso el perceptrón, una estructura más flexible compuesta por pesos ajustables que permitían “aprender” relaciones lineales entre datos. Su idea, revolucionaria en su época, abrió la puerta a los sistemas adaptativos. Sin embargo, las limitaciones del perceptrón se hicieron evidentes pronto: no podía capturar relaciones no lineales, y su potencial parecía, en ese momento, estrecho. Durante un tiempo, esto produjo un estancamiento conceptual conocido como el “invierno de las redes neuronales”.

El renacimiento llegó cuando se introdujo el algoritmo de retropropagación del error en la década de 1980 (Rumelhart et al., 1986), que permitió entrenar redes de múlti-

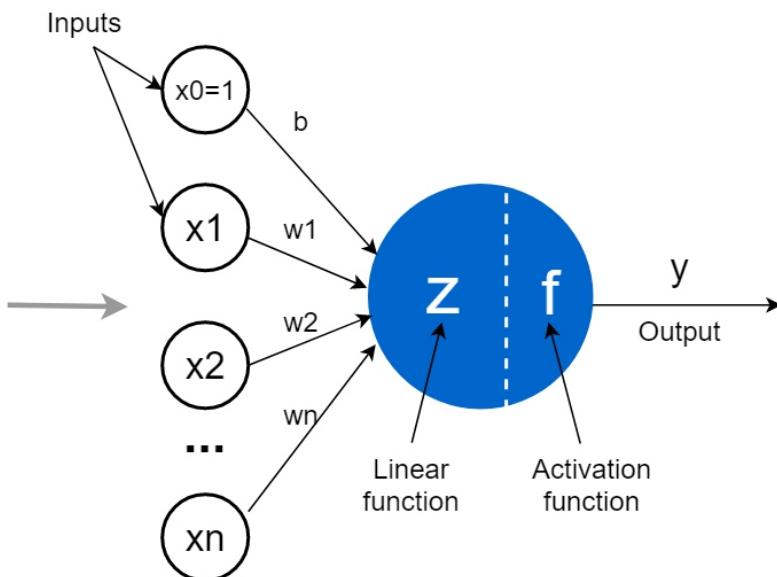
iples capas. Este avance transformó el campo por completo. De repente, los modelos podían resolver problemas complejos, ajustar millones de parámetros y descubrir representaciones internas sin supervisión explícita. La presentación Inteligencias Paralelas retoma ese hilo histórico y lo conecta con ejemplos contemporáneos de aprendizaje automático, mostrando cómo ese salto conceptual permitió el desarrollo de redes profundas utilizadas en visión por computadora, reconocimiento de lenguaje y análisis de grandes volúmenes de datos. Hoy, arquitecturas como las redes convolucionales, las redes recurrentes, los transformadores y los modelos generativos constituyen la columna vertebral de muchos sistemas de inteligencia artificial.

Aunque la terminología toma prestados conceptos de la neurobiología, los elementos que conforman una red artificial pertenecen al dominio estricto de las matemáticas y la estadística. Una unidad artificial recibe entradas numéricas, las multiplica por pesos, suma un sesgo y aplica una función de activación —una operación aparentemente trivial, pero fundamental para introducir no linealidad—. Esta combinación, repetida millones de veces, es la responsable de la riqueza expresiva de los modelos contemporáneos (Goodfellow et al., 2016). Ver Figura 2.

Los pesos representan la importancia relativa de cada señal de entrada. El sesgo permite desplazar la función de activación, modificando su sensibilidad. Finalmente, la activación determina la transformación aplicada: puede ser sigmoidal, hipérbólica, lineal o basada en unidades rectificadas como ReLU, ampliamente utilizadas por su eficiencia computacional.

Cuando varias de estas unidades se organizan en capas, se forma una arquitectura que procesa información de manera jerárquica. Cada capa detecta patrones de dife-

Figura 2. Diagrama de un perceptrón artificial. Adaptado R. Pramoditha (2021)



Nota: perceptrón artificial. Adaptado de The concept of artificial neurons (perceptrons) in neural networks por R. Pramoditha (2021), Towards Data Science. <https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc>

rente complejidad: las iniciales captan características simples; las intermedias, patrones abstractos; las finales, decisiones de mayor orden.

Así funciona el procesamiento en redes convolucionales aplicadas a visión por computador, donde filtros sucesivos extraen gradualmente bordes, contornos, formas y finalmente objetos completos.

El aprendizaje en redes artificiales no se basa en mecanismos electroquímicos ni en sinapsis moduladas por neurotransmisores. En cambio, se fundamenta en la optimización matemática. El objetivo consiste en minimizar una función de pérdida que mide qué tan lejos está el modelo de la predicción correcta. Para lograrlo, se emplea el descenso del gradiente y sus variantes, que ajustan los pesos de forma iterativa a partir del error.

Gracias al algoritmo de retropropagación, las redes aprenden calculando cómo cada parámetro contribuyó al error total y actualizando su valor para mejorar el desempeño en el siguiente ciclo. Aunque este procedimiento se repite millones de veces, ocurre a velocidades extraordinarias gracias a las unidades de procesamiento paralelo como las GPU y, más recientemente, las TPU.

A diferencia del cerebro, que aprende incluso con pocos ejemplos y en entornos ambiguos, las redes artificiales suelen requerir grandes vo-

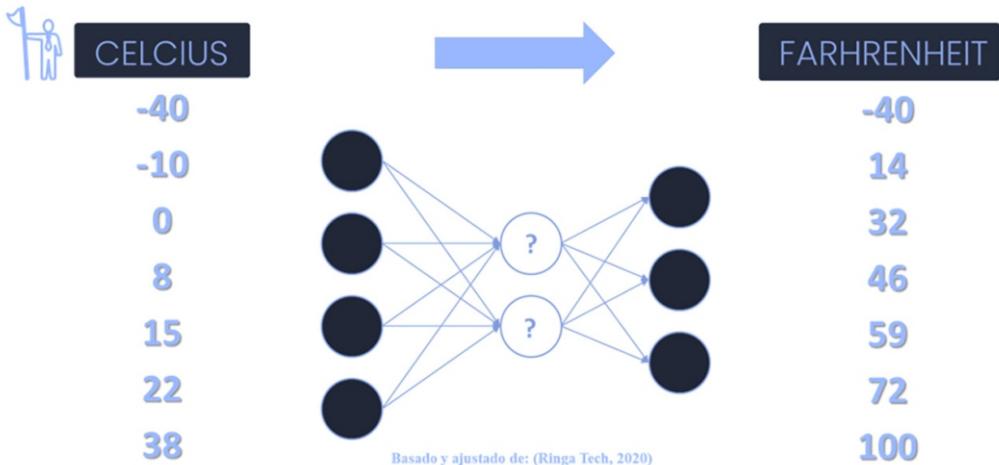
lúmenes de datos para obtener un rendimiento aceptable. Esta diferencia no se debe a incapacidad computacional, sino a que los modelos matemáticos carecen de los mecanismos biológicos de generalización, contexto y experiencia acumulada. El cerebro, como señalan Kandel et al. (2013), construye conocimiento sobre múltiples escalas temporales; las redes artificiales, en cambio, aprenden estadísticamente sobre representaciones distribuidas.

Un ejemplo sencillo pero poderoso: entrenar una neurona para convertir grados Celsius a Fahrenheit. Aunque la relación es lineal y no requiere de redes profundas, este caso demuestra la esencia del aprendizaje automático: el modelo no sabe la fórmula inicial, pero la descubre a través de ejemplos repetidos.

La conversión entre grados Celsius y Fahrenheit permite observar con claridad cómo difieren dos formas de resolver un mismo problema: la programación tradicional y el aprendizaje automático. En un enfoque clásico, el programador conoce la fórmula exacta ($F = 1.8C + 32$) y la codifica paso a paso.

El computador no descubre nada; simplemente ejecuta instrucciones precisas que garantizan un resultado determinista. Es la esencia del paradigma algorítmico: reglas explícitas, operaciones definidas y ausencia total de incertidumbre. Ver Figura 3.

Figura 3. Adaptación de (Ringa Tech, 2020)



Nota: Adaptación propia del autor para presentación.

En contraste, un modelo de inteligencia artificial no recibe la fórmula. Aprende a partir de ejemplos. La red neuronal observa pares de datos —por ejemplo, 10°C corresponde a 50°F— y ajusta sus pesos internos para aproximarse cada vez más a la respuesta correcta.

Este proceso de optimización mediante retropropagación, descrito originalmente por Rumelhart et al. (1986), permite que el modelo infiera la relación lineal sin haberla visto explícitamente. La presentación Inteligencias Paralelas muestra justamente cómo la red converge hacia la función correcta sin instrucciones directas, solo gracias a patrones en los datos.

El contraste es profundo: en la computación tradicional el conocimiento está en el código; en los mo-

delos de IA, está en los datos. La máquina deja de ejecutar instrucciones y comienza a identificar regularidades, un enfoque que recuerda —de forma lejana, pero útil— la manera en que los sistemas biológicos aprenden a partir de experiencia (Kandel et al., 2013).

Los modelos modernos han evolucionado hacia estructuras colosales con capacidades sorprendentes. Las redes convolucionales son especializadas en análisis de imágenes; las recurrentes se orientan al procesamiento de secuencias; los transformadores han revolucionado el campo del procesamiento del lenguaje natural gracias al mecanismo de atención, que permite identificar qué partes de una entrada son más relevantes para una tarea específica (Vaswani et al., 2017).

Esta explosión de arquitecturas ha estado acompañada por avances en hardware y disponibilidad de datos. Pero también ha reabierto el debate sobre la naturaleza de la inteligencia: ¿son estos modelos realmente “inteligentes” o simplemente potentes optimizadores estadísticos? Investigadores como LeCun et al. (2015) muestran que la capacidad de estos modelos para captar patrones sutiles supera con creces lo que se consideraba posible hace apenas una década. No obstante, su tipo de “comprensión” dista profundamente de la cognición biológica.

Por muy avanzadas que sean, estas redes siguen dependiendo de estructuras matemáticas simples que se repiten a gran escala. Su aprendizaje es eficiente, pero estrecho; su razonamiento es profundo, pero sin comprensión semántica humana. Aun así, su valor práctico es incuestionable.

Puentes y Contrastes entre Redes Biológicas y Redes Artificiales

El diálogo entre neurociencia e inteligencia artificial se sostiene sobre un delicado equilibrio entre inspiración y distanciamiento. Las redes neuronales artificiales surgieron como una metáfora de las biológicas, pero con el tiempo evolucionaron hasta convertirse en sistemas radicalmente distintos. La neurociencia, por su parte, encon-

tró en la IA un marco conceptual útil para imaginar cómo podrían organizarse ciertos procesos cognitivos; sin embargo, la evidencia ha demostrado que las semejanzas entre ambos sistemas son más metafóricas que estructurales. Examinar estos puentes y divergencias no solo aclara malentendidos comunes, sino que también ilumina lo que cada disciplina puede aprender de la otra.

La inspiración inicial de las redes neuronales artificiales fue explícitamente biológica. McCulloch y Pitts (1943) imaginaron una neurona como una unidad lógica capaz de integrar señales y producir una salida binaria. Décadas más tarde, el perceptrón de Rosenblatt (1958) reforzó esa intuición con mecanismos de ajuste de pesos que recorrián, superficialmente, a la plasticidad sináptica. Pero el parecido terminaba ahí.

Mientras que una neurona biológica integra procesos electroquímicos que involucran neurotransmisores, modulación glial y cascadas metabólicas (Kandel et al., 2013), una “neurona artificial” es simplemente una función matemática que combina entradas y aplica una transformación no lineal. No posee morfología, variabilidad química ni dinámica temporal compleja. Es una abstracción. Una simplificación extrema que, sin embargo, ha sido extraordinariamente útil para desarrollar sistemas de aprendizaje.

Este tipo de inspiración distante no pretende replicar el cerebro, sino capturar algunos principios funcionales: procesamiento distribuido, ajuste adaptativo de conexiones, jerarquías de información. Como señalan LeCun et al. (2015), las redes artificiales “se parecen al cerebro tanto como un avión se parece a un ave”: comparten conceptos, no estructuras.

A pesar de la distancia estructural, existen paralelos conceptuales que facilitan la comprensión de ambos sistemas. La plasticidad sináptica —ya sea la potenciación a largo plazo o la depresión sináptica— puede compararse, en términos muy generales, con el ajuste de pesos en una red artificial. Ambos procesos modifican la fuerza de conexiones para mejorar el desempeño en una tarea específica.

Del mismo modo, la modularidad del cerebro, observada en su conectoma funcional, encuentra una analogía en las capas jerárquicas de las redes profundas (Bassett & Sporns, 2017). En ambos casos, distintos “módulos” se especializan en ciertos tipos de procesamiento y luego combinan su actividad para producir una salida coherente.

Estas analogías, sin embargo, deben manejarse con cautela. Una sinapsis no es un peso; una neurona no es una función de activación; una red biológica no es una arquitectura matemática. Las comparaciones funcionan como puentes pe-

dagógicos, útiles para explicar conceptos complejos, pero insuficientes para describir las realidades profundas de cada sistema.

La divergencia entre redes biológicas y artificiales es abismal cuando se observan con detalle. El cerebro opera con una eficiencia energética extraordinaria —solo 20 vatios para sostener miles de millones de operaciones simultáneas—, mientras que los grandes modelos de IA pueden requerir megavatios de energía para entrenarse (Faisal et al., 2008). El contraste no es solo cuantitativo, sino cualitativo.

Las redes biológicas funcionan en un entorno ruidoso, variable y estocástico. Este ruido, lejos de ser un problema, facilita la adaptabilidad y la capacidad de generalización. En cambio, las redes artificiales dependen de estabilidad, precisión numérica y entornos altamente controlados. El cerebro reorganiza su conectividad en escalas temporales que van de milisegundos a meses; los modelos artificiales ajustan sus parámetros en microsegundos, pero carecen de la rica dimensión temporal que caracteriza la cognición humana.

Además, las redes artificiales aprenden patrones estadísticos sin entender su significado. Los modelos pueden identificar correlaciones extremadamente complejas, pero no poseen conciencia, intención ni comprensión contextual. La cognición humana, como señalan Kan-

del et al. (2013), está profundamente influenciada por factores emocionales, motivacionales y sociales que no tienen equivalente en la IA contemporánea.

El aprendizaje humano es un proceso multimodal que integra percepción, emoción, memoria, motivación y experiencia previa. No consiste simplemente en reconocer patrones, sino en atribuirles significado. Aprender implica interpretar el mundo a través de lentes biográficas y culturales. La IA, en cambio, aprende ajustando parámetros para minimizar errores en una tarea específica. Sus representaciones internas —por muy sofisticadas que sean— no tienen contenido semántico propio.

Este contraste se evidencia en el hecho de que los seres humanos pueden aprender con pocos ejemplos gracias a mecanismos de abstracción profunda, mientras que los modelos artificiales dependen, en general, de volúmenes enormes de datos etiquetados (Lake et al., 2017). El cerebro puede generalizar desde un caso único; las redes artificiales suelen necesitar miles.

Sin embargo, el aporte conceptual es recíproco. La IA ha permitido generar teorías sobre cómo podrían representarse ciertos cálculos cognitivos, mientras que la neurociencia ha inspirado nuevas arquitecturas. Son campos que avanzan paralelos, no en competencia, sino en diálogo.

Perspectivas de Convergencia: Neuroingeniería, IA y Cerebro

La intersección entre neurociencia e inteligencia artificial no solo revela diferencias profundas entre ambos sistemas, sino también puntos de aproximación que están abriendo horizontes tecnológicos completamente nuevos. En décadas recientes, los avances en interfaces cerebro-computador, computación neuromórfica y modelos híbridos han permitido vislumbrar escenarios donde las fronteras tradicionales entre lo biológico y lo artificial se vuelven más permeables. Esta convergencia no busca imitar al cerebro ni reemplazarlo, sino estudiar cómo sus principios pueden inspirar tecnologías más eficientes, flexibles y adaptativas. A la vez, los modelos computacionales ofrecen herramientas para reinterpretar funciones del sistema nervioso desde nuevas perspectivas.

Las interfaces cerebro-computador (BCI por sus siglas en inglés) son uno de los campos donde la colaboración entre neurociencia e ingeniería ha sido más fructífera. Estos sistemas permiten traducir la actividad neuronal en comandos para dispositivos externos, posibilitando, por ejemplo, que personas con parálisis puedan mover prótesis robóticas o comunicarse mediante patrones de neuroactividad (Serruya et al., 2002). El principio es simple en concepto, pero complejo en ejecución: registrar señales cerebrales mediante electrodos invasi-

vos o no invasivos, procesarlas con algoritmos de decodificación y convertirlas en acciones digitales.

El conocimiento profundo sobre la organización cortical y las rutas motoras ha permitido desarrollar modelos de decodificación capaces de inferir intenciones de movimiento a partir de patrones neuronales. Investigaciones recientes muestran, incluso, la posibilidad de restaurar funciones sensoriomotoras mediante ciclos cerrados de retroalimentación, donde la señal decodificada genera un estímulo háptico que regresa al usuario, cerrando un circuito entre el cerebro y la máquina (Shenoy & Carmena, 2014).

Los avances en BCI plantean preguntas fascinantes: ¿puede una máquina extender las capacidades humanas? ¿Es posible integrar señales artificiales en el cerebro sin alterar su integridad funcional? Aunque estas tecnologías aún están en desarrollo, su potencial terapéutico y rehabilitador es extraordinario.

Por su parte, la computación neuromórfica surge del deseo de construir sistemas que no solo se inspiren conceptualmente en el cerebro, sino que también adopten algunos de sus principios estructurales. Chips como *TrueNorth* de IBM o *Loihi* de Intel implementan neuronas de picos (*spiking neurons*) que transmiten información en forma de eventos discretos, imitando la diná-

mica temporal de las neuronas biológicas (Davies et al., 2018). Estos dispositivos consumen muy poca energía y permiten ejecutar modelos con eficiencia notable, especialmente en tareas de reconocimiento sensorial y control robótico.

A diferencia de las redes tradicionales, las *spiking neural networks* (SNN) integran el tiempo como una variable fundamental, aproximándose a la forma en que las neuronas reales procesan señales. Aunque todavía están en desarrollo, representan un puente prometedor entre las limitaciones del hardware tradicional y las propiedades emergentes del cerebro humano.

El estudio de la arquitectura neural, especialmente del conectoma, ha inspirado también modelos de procesamiento distribuido que imitan la organización modular del cerebro (Bassett & Sporns, 2017). Estos modelos no buscan replicar estructuras biológicas, sino aprovechar principios como la eficiencia energética, la conectividad jerárquica y la descentralización del procesamiento.

La convergencia entre biología e inteligencia artificial abre caminos potentes, pero también inquietudes profundas. Las BCI plantean desafíos sobre privacidad neuronal: si es posible decodificar patrones de intención, ¿qué salvaguardas se requieren para evitar su uso indebido? Investigadores como Yuste et al. (2017) han propuesto marcos

éticos para proteger lo que llaman “neuroderechos”, incluyendo el derecho a la identidad mental, la privacidad de la actividad cerebral y la libertad cognitiva.

En paralelo, la IA plantea dilemas sobre autonomía, responsabilidad y agencia. Los modelos actuales no comprenden el sentido de sus decisiones, pero sus resultados pueden influir en sistemas críticos como diagnóstico médico, justicia o educación. La interacción entre ambos campos exige una reflexión ética profunda sobre cómo diseñar tecnologías que amplíen la capacidad humana sin erosionar la dignidad y la autonomía.

La filosofía de la mente también encuentra un espacio en esta convergencia. Preguntas como “¿puede una máquina tener experiencias?” o “¿es la inteligencia reducible a funciones computacionales?” retoman vigor ante avances tecnológicos que, aunque no replican la cognición humana, sí la emulan en ciertos comportamientos superficiales. La diferencia ontológica entre un modelo estadístico y un cerebro consciente sigue siendo uno de los desafíos intelectuales más importantes de nuestra época.

Entre las tendencias emergentes en inteligencia artificial, pocas resultan tan conceptualmente sugerentes como el *machine unlearning*. Este campo busca permitir que los modelos eliminen información previamente aprendida sin ne-

cesidad de reconstruirse desde cero. En otras palabras, intenta que un sistema artificial “olvide” ciertos datos o patrones, ya sea para corregir sesgos, respetar solicitudes de privacidad (como el derecho al olvido) o depurar representaciones internas no deseadas (Cao & Yang, 2015; Xu et al., 2023).

Aunque el mecanismo es estrictamente computacional, guarda ecos curiosos con el funcionamiento del cerebro. El olvido biológico no es un fallo; es un mecanismo adaptativo esencial. La poda sináptica durante el desarrollo y los procesos de depresión a largo plazo (LTD por sus siglas en inglés) permiten eliminar conexiones innecesarias para mejorar la eficiencia y evitar la saturación cognitiva (Kandel et al., 2013). En el aprendizaje humano, olvidar es tan importante como recordar.

Los métodos computacionales actuales emplean técnicas como la fragmentación de datos, el reentrenamiento diferencial y la actualización localizada de parámetros para suprimir la influencia de ejemplos específicos en el modelo. En esencia, se busca revertir el rastro estadístico que ciertos datos dejaron durante el entrenamiento, imitando —de forma muy lejana, pero conceptualmente inspiradora— la capacidad del cerebro para debilitar o eliminar conexiones sinápticas.

El paralelismo no implica equivalencia. El olvido biológico es orgá-

nico, contextual, influido por emociones y experiencias. El desaprendizaje artificial es algorítmico, limitado y cuantificable. Pero ambos comparten una idea central: la eficiencia cognitiva requiere no solo adquirir información, sino también saber qué dejar atrás.

Más Allá de los Cables y las Sinapsis: Reflexiones Finales sobre Dos Inteligencias que Convergen

La comparación entre redes neuronales biológicas y redes neuronales artificiales no busca resolver un debate de superioridad, sino abrir una ventana hacia la complejidad de dos sistemas que, desde dominios radicalmente distintos, intentan responder a la misma pregunta: ¿cómo surge la inteligencia? A lo largo de este recorrido queda claro que ambos modelos —uno moldeado por millones de años de evolución y el otro construido mediante abstracciones matemáticas— iluminan diferentes aspectos del procesamiento de información. Sus coincidencias son conceptuales; sus diferencias, profundas. Sin embargo, es precisamente en ese espacio intermedio donde emergen las oportunidades más ricas.

El cerebro humano, con su plasticidad, su economía energética y su capacidad para generar significado, continúa siendo una fuente inagotable de inspiración para la ingeniería. La evidencia en neurociencia muestra que el aprendizaje no

es solo un proceso de fortalecimiento sináptico, sino un equilibrio entre creación y eliminación, un sistema dinámico que redefine continuamente sus rutas internas (Kandel et al., 2013). Esa flexibilidad, difícil de capturar en términos computacionales, es esencial para explicar cómo los seres humanos pueden aprender con pocos ejemplos, adaptarse a entornos inciertos y reorganizar funciones tras una lesión.

Por otro lado, los modelos de inteligencia artificial han alcanzado niveles de desempeño que superan con creces la capacidad humana en tareas específicas. Sistemas capaces de analizar imágenes con precisión milimétrica o sintetizar lenguaje con fluidez sorprendente revelan que la eficiencia matemática puede, en ciertos dominios, compensar la carencia de comprensión semántica profunda (LeCun et al., 2015). Sin embargo, los modelos artificiales continúan siendo, en esencia, estructuras de optimización: aprenden correlaciones, no significados. Sus representaciones internas carecen de la dimensión subjetiva que caracteriza la cognición humana.

El diálogo entre ambas disciplinas no implica que la inteligencia artificial replique al cerebro, ni que el cerebro funcione como un algoritmo. Más bien, su interacción genera un intercambio fecundo: la neurociencia aporta principios como la modularidad, la plasticidad y la eficien-

cia energética; la IA ofrece modelos formales para explorar hipótesis y herramientas para interpretar patrones que antes eran invisibles. En campos como la computación neuromórfica, las interfaces cerebro-computador y el machine unlearning, esta convergencia es especialmente evidente, señalando un futuro donde lo biológico y lo artificial pueden complementarse de maneras cada vez más sofisticadas (Davies et al., 2018; Xu et al., 2023).

El futuro de esta intersección dependerá de dos factores clave. Por un lado, de la capacidad para desarrollar tecnologías que respeten la complejidad del cerebro sin caer en reduccionismos simplistas. Por otro, de la responsabilidad ética para garantizar que los sistemas artificiales amplíen, y no limiten, las capacidades humanas. Conceptos emergentes como los neuroderechos, la auditabilidad de modelos y el diseño seguro de sistemas híbridos serán esenciales para asegurar una relación equilibrada entre ambas inteligencias (Yuste et al., 2017).

En última instancia, pensar el cerebro junto con la inteligencia artificial no transforma únicamente nuestra forma de diseñar tecnología. También transforma la manera en que entendemos lo que significa aprender, recordar, olvidar y adaptarnos. Entre sinapsis que se fortalecen y algoritmos que se optimizan, surge un terreno conceptual donde las

preguntas más antiguas de la filosofía —sobre la mente, el conocimiento y la conciencia— encuentran nuevas formas de expresarse. Allí, en el encuentro entre cables y sinapsis, entre señales eléctricas y datos, se dibuja el contorno de una inteligencia que es, al mismo tiempo, humana y artificial.

Referencias

- Bassett, D. S., & Sporns, O. (2017). *Network neuroscience*. *Nature Neuroscience*, 20(3), 353–364. <https://doi.org/10.1038/nn.4502>
- Bliss, T. V. P., & Lømo, T. (1973). *Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path*. *Journal of Physiology*, 232(2), 331–356.
- Cao, Y., & Yang, J. (2015). *Towards making systems forget with machine unlearning*. In 2015 IEEE Symposium on Security and Privacy (pp. 463–480). <https://doi.org/10.1109/SP.2015.35>
- Davies, M., Srinivasa, N., Lin, T. H., Chinya, G., et al. (2018). *Loihi: A neuromorphic manycore processor with on-chip learning*. *IEEE Micro*, 38(1), 82–99.
- Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection*. Basic Books.
- Edlow, B. L., & Menon, D. K. (2024). *Brain network connectivity and consciousness*. *Critical Care Medicine*, 52(9), 1414–1426.
- Faisal, A. A., Selen, L. P., & Wolpert, D. M. (2008). *Noise in the nervous*

- system. *Nature Reviews Neuroscience*, 9(4), 292–303.
- Fox, M. D. (2018). *Mapping human brain networks*. New England Journal of Medicine, 379(23), 2237–2245.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Isaacson, J. S., & Scanziani, M. (2011). *How inhibition shapes cortical activity*. *Neuron*, 72(2), 231–243.
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (2013). *Principles of neural science* (5th ed.). McGraw-Hill.
- Lake, B. M., Ullman, T., Tenenbaum, J. B., & Gershman, S. (2017). *Building machines that learn and think like people*. *Behavioral and Brain Sciences*, 40, e253.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- McCulloch, W. S., & Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. *Bulletin of Mathematical Biophysics*, 5, 115–133. <https://doi.org/10.1007/BF02478259>
- Mountcastle, V. (1997). *The columnar organization of the neocortex*. *Brain*, 120(4), 701–722.
- Nayak, M., et al. (2022). *Regional neuronal plasticity*. *Heliyon*, 8(12), e12292.
- Rosenblatt, F. (1958). *The perceptron: A probabilistic model for information storage and organization in the brain*. *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning representations by back-propagating errors*. *Nature*, 323(6088), 533–536.
- Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R., & Donoghue, J. P. (2002). *Instant neural control of a movement signal*. *Nature*, 416(6877), 141–142.
- Shenoy, K. V., & Carmena, J. M. (2014). *Combining decoder design and neural adaptation in brain–machine interfaces*. *Neuron*, 84(4), 665–680.
- Sporns, O. (2011). *Networks of the brain*. MIT Press.
- Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). *Attention is all you need*. In *Advances in Neural Information Processing Systems* (Vol. 30).
- Xu, X., Ren, K., Yu, Z., & Zhang, S. (2023). *Machine unlearning: A survey*. *ACM Computing Surveys*, 55(8), 1–37. <https://doi.org/10.1145/3527448>
- Yuste, R., et al. (2017). *Four ethical priorities for neurotechnologies and AI*. *Nature*, 551(7679), 159–163. 

Joshsua J. González Díaz es Ingeniero de Sistemas de la Pontificia Universidad Javeriana, Especialista en Seguridad de la Información de la Universidad de los Andes, Especialista en Derecho Informático y de las Nuevas Tecnologías de la Universidad Externado de Colombia, en conjunto con la Universidad Complutense de Madrid. Es Magíster en Seguridad de la Información de la Universidad de los Andes y Magíster en Derecho Informático y de las Nuevas Tecnologías del Externado. Se ha formado también como Gestor de Ciberseguridad para Gerentes en el MIT Sloan School of Management y actualmente es candidato a Doctor (PhD(c)) en Cybersecurity Analytics and AI en la George Washington University. Se desempeña como Chief Information Security Officer (CISO), Data Privacy Officer (DPO) y Chief Compliance Officer (CCO) en una empresa internacional, liderando la estrategia de ciberseguridad, cumplimiento normativo y privacidad de datos. Cuenta con certificaciones internacionales como CEH, CHFI, ECSA, LPT, ISO/IEC 27001:2022 Lead Auditor, ISO 27032:2018 Cybersecurity Lead Manager, ISO 31000:2018 Risk Management e Incident Response and Forensics (CIRF), entre otras, que respaldan su amplia experiencia en gobierno, gestión y respuesta a incidentes de seguridad de la información.