CAPÍTULO V

Panarquía artificial: ciclos adaptativos de la IA

David Martinez Rego*

Este trabajo propone un marco analítico para comprender la evolución reciente de la inteligencia artificial (IA), superando la narrativa simplista de "veranos e inviernos". A través de la teoría de ciclos adaptativos y panarquía, se interpreta la IA como un sistema sociotécnico complejo, donde invención, innovación y difusión interactúan en múltiples escalas. Se revisan hitos históricos, desde la teoría estadística hasta el aprendizaje profundo y los grandes modelos de lenguaje, y se analizan limitaciones, anomalías empíricas y la transición hacia aplicaciones prácticas. El enfoque multinivel permite reconciliar expectativas extremas con la realidad, ofreciendo una visión prospectiva para la gestión estratégica de la tecnología.

Palabras clave: inteligencia artificial, panarquía, ciclos adaptativos, teoría del aprendizaje, aprendizaje profundo.

* Agradecemos a Funcas la invitación al evento y la oportunidad de participar en la presente publicación.

1. INTRODUCCIÓN

El debate contemporáneo en torno a la inteligencia artificial (IA) se caracteriza por una tensión constante entre entusiasmo desmedido y preocupación social, alimentada en gran medida por discursos mediáticos, intereses económicos y posicionamientos geopolíticos. En este contexto, resulta cada vez más difícil construir marcos analíticos que permitan comprender el estado real de la tecnología, sus límites estructurales, sus capacidades emergentes y su evolución futura desde una perspectiva científicamente fundamentada.

El objetivo de este trabajo es proponer un marco interpretativo para analizar la evolución reciente de la inteligencia artificial, alejándose del modelo narrativo dominante de "veranos e inviernos de la IA". Se argumenta que dicho modelo, aunque intuitivo, resulta insuficiente y potencialmente engañoso para comprender dinámicas complejas y acumulativas que caracterizan el desarrollo actual de la disciplina.

Como alternativa, se introduce un enfoque inspirado en la teoría de sistemas complejos, concretamente los conceptos de ciclo adaptativo y panarquía, originalmente desarrollados en el ámbito de la ecología por Holling y colaboradores (Holling, 2001). Este marco permite analizar la IA como un componente integrado en un ecosistema sociotécnico, sujeto a dinámicas de conservación, ruptura, reorganización y explotación que operan simultáneamente en múltiples escalas.

Es fundamental reconocer que el marco de comprensión que adoptemos condiciona no solo nuestras proyecciones mentales sobre el futuro de la inteligencia artificial, sino también nuestras decisiones colectivas frente a cambios tecnológicos. La teoría de sistemas enfatiza que un análisis incorrecto de un sistema complejo puede conducir a decisiones sociales y políticas indeseables, generando consecuencias contraproducentes a nivel económico, institucional o ético (Meadows and Wright, 2008). En el caso de la IA, la falta de un marco sólido de interpretación puede amplificar sesgos, malentendidos o expectativas erróneas, afectando la manera en que se implementan, regulan y adoptan estas tecnologías. Por ello, estudiar y formalizar un marco analítico riguroso para la IA no es solo un ejercicio académico, sino una necesidad estratégica para guiar la interacción de la sociedad con estas herramientas emergentes de manera informada, responsable y sostenible.

2. EL MODELO DE "VERANOS E INVIERNOS" DE LA IA

La narrativa de los "veranos e inviernos" ha sido ampliamente utilizada tanto en divulgación como en literatura académica para describir periodos alternantes de avances espectaculares y estancamiento de una tecnología. Según este modelo, las expectativas generadas por nuevos enfoques técnicos conducen a ciclos de inversión excesiva, seguidos inevitablemente por desilusión cuando dichas expectativas no se materializan.

2.1. Origen y difusión del modelo

Un ejemplo paradigmático de este ciclo corresponde al periodo comprendido entre finales de la década de 1980 y principios de los 2000. Hacia finales de los años 80 se produjo lo que podría considerarse un *verano* de la IA, caracterizado por un entusiasmo renovado en torno a los modelos neuronales y su potencial para la generalización, sustentado por avances en la retropropagación y en redes convolucionales (Rumelhart *et al.*, 1986; LeCun *et al.*, 1989). Sin embargo, la experiencia práctica pronto mostró que estos modelos carecían de la capacidad de generalizar de manera amplia y eficiente, limitando su aplicabilidad más allá de casos específicos. Esta situación llevó a un *invierno* prolongado, que se extendió aproximadamente hasta principios de la década del 2000, durante el cual la investigación en redes neuronales perdió prominencia frente a enfoques alternativos, como los sistemas expertos y los métodos de aprendizaje supervisado basados en teoría estadística (Mitchell, 1997). La reflexión crítica de muchos expertos en esta época es feroz y subrayaba que la imitación literal de procesos biológicos, aunque instructiva, no necesariamente conducía al desarrollo de sistemas de aprendizaje artificial efectivos[1]. Este periodo evidencia que las expectativas generadas durante los 80 con los modelos neuronales no estaban fundadas sobre ninguna base sólida, resaltando la importancia de comprender los límites teóricos y las condiciones de aplicabilidad de cada enfoque de aprendizaje.

Reacción a las limitaciones de los modelos neuronales

Las limitaciones prácticas encontradas por los modelos neuronales a finales de la década de 1980 impulsaron el surgimiento de avances conceptuales que hoy constituyen los fundamentos del aprendizaje automático moderno. Por un lado, la incapacidad de las redes neuronales de generalizar más allá de conjuntos de datos relativamente pequeños motivó el desarrollo de la teoría del aprendizaje estadístico. Por otro lado, la constatación de que ningún algoritmo de aprendizaje podría resolver de manera óptima todos los problemas posibles condujo a la formulación del teorema de *No Free Lunch* (NFL), que estableció límites universales a la optimización algorítmica (Wolpert and Macready, 1997).

Teoría del aprendizaje estadístico

El desarrollo del aprendizaje automático durante las décadas de 1980 y 1990 estuvo fuertemente influido por la teoría del aprendizaje estadístico, formalizada principalmente por Vapnik y Chervonenkis (Vapnik, 1998). Uno de sus resultados centrales establece que el error de generalización de un modelo depende de su complejidad (capacidad) y del tamaño del conjunto de datos de entrenamiento (ver figura 1).

[1] Vapnik (1998) llega a comparar la aproximación con inspirarse en el vuelo de las aves para la construcción de aviones.

FIGURA 1

PRINCIPIO FUNDAMENTAL DEL APRENDIZAJE ESTADÍSTICO ESTABLECE QUE EL ERROR DE GENERALIZACIÓN DE UN ALGORITMO A SOBRE UN PROBLEMA P ESTÁ ACOTADO SUPERIORMENTE POR LA SUMA DEL ERROR OBSERVADO EN EL CONJUNTO DE ENTRENAMIENTO D Y UN TÉRMINO PROPORCIONAL A LA RELACIÓN ENTRE LA COMPLEJIDAD DEL ALGORITMO A Y EL TAMAÑO DEL CONJUNTO D. DURANTE EL "INVIERNO" ORIGINAL DE LA IA EN LA DÉCADA DE 1990, LOS MODELOS NEURONALES FUERON CONSIDERADOS EXCESIVAMENTE COMPLEJOS EN RELACIÓN CON LOS CONJUNTOS DE DATOS DISPONIBLES, LO QUE SUGERÍA QUE SU ERROR EN LA PRÁCTICA NO PODÍA SER CONTROLADO, EXPLICANDO ASÍ SU LIMITADA ADOPCIÓN EN APLICACIONES REALES

$$\text{Error}(A,P) \leq \text{Error}_{\text{Entrenamiento}}(A,D) + \frac{\text{Complejidad}(A)}{\text{Tamaño}(D)}$$

De manera simplificada, puede afirmarse que modelos con mayor capacidad requieren conjuntos de datos proporcionalmente mayores para evitar el sobreajuste y un mal funcionamiento en la práctica. Este principio influyó profundamente en la percepción de las redes neuronales, consideradas excesivamente complejas para los recursos computacionales y los datos disponibles en ese momento.

El teorema de No Free Lunch

El teorema de *No Free Lunch* (NFL), formulado por Wolpert y Macready (1997), reforzó este escepticismo al demostrar que no existe un algoritmo de optimización universalmente superior cuando se promedian todos los posibles problemas. Este resultado fue interpretado como una limitación fundamental a la idea de algoritmos generales de aprendizaje, favoreciendo enfoques altamente especializados.

Efectos académicos y políticos del modelo

Estos dos desarrollos teóricos no solo fueron adoptados como explicación al estancamiento temporal de las redes neuronales durante la época, sino que ofrecían un certificado teórico de que la aproximación carecía de fundamento práctico, pues la cantidad de datos y cómputo necesarios para asegurar su funcionamiento práctico nunca iban a estar disponibles y su admisión como principio universal de aprendizaje contradice el principio de *No Free Lunch*.

Como consecuencia, tanto la academia como la industria aceptaron de manera generalizada el *statu quo* impuesto por las limitaciones de los modelos neuronales y redujeron drásticamente la financiación destinada a esta línea de investigación. La mayoría

de los investigadores en redes neuronales abandonaron progresivamente este enfoque, desplazándose hacia el desarrollo de modelos especializados de aprendizaje supervisado que pudieran garantizar resultados prácticos en aplicaciones concretas (Mitchell, 1997). Este cambio de trayectoria consolidó un paradigma centrado en modelos de capacidad limitada y conjuntos de datos relativamente pequeños, fomentando la investigación aplicada en problemas específicos y ralentizando el progreso en el desarrollo de arquitecturas neuronales.

3. EL RESURGIMIENTO EMPÍRICO DEL APRENDIZAJE PROFUNDO

El resurgimiento del aprendizaje profundo en la última década se ha manifestado de manera particularmente visible a través de dos fenómenos tecnológicos y sociales: la eficacia de los modelos de procesamiento de imágenes y la irrupción de los grandes modelos de lenguaje (más conocidos por la siglas de su nombre anglosajón *Large Language Models* o los LLM) en aplicaciones de interacción natural con la información textual. El punto de inflexión definitivo se da en el año 2022, cuando el uso masivo de ChatGPT da visibilidad generalizada de los LLM en la sociedad.

Desde la perspectiva tradicional de los ciclos de "veranos e inviernos" de la IA, este fenómeno se ha interpretado de dos maneras contrapuestas. Por un lado, existe una corriente que mantiene que la teoría del aprendizaje estadístico sigue siendo válida y robusta, y que los avances actuales no hacen sino evidenciar los límites que tarde o temprano aparecerán nuevamente, anticipando un futuro "invierno" de la IA. Por otro lado, hay quienes consideran que los inviernos históricos han quedado atrás y que los principios teóricos eran, en el mejor de los casos, irrelevantes o incluso incorrectos para comprender la dinámica de los modelos neuronales modernos.

Como alternativa a estas interpretaciones simplistas, sostenemos que este hito no representa un cambio abrupto en la teoría, sino el resultado de un proceso acumulativo de más de dos décadas. Durante este tiempo, los principios de diseño derivados de la teoría estadística han permanecido presentes, guiando tanto la construcción de modelos como la organización de conjuntos de datos y el desarrollo de la industria de procesamiento de información. La expansión de tecnologías para la captura y procesamiento masivo de datos permitió la observación de fenómenos empíricos que antes eran inaccesibles, revelando límites prácticos de la teoría y comportamientos de aprendizaje que, aunque no anticipados, resultan reproducibles en distintos dominios y tareas. Estos resultados empíricos, obtenidos bajo condiciones de datos masivos y cómputo distribuido, permiten extender la teoría del aprendizaje estadístico: muestran cómo las estructuras internas de los modelos neuronales pueden alinearse con tareas generales, producir generalización efectiva y aprender representaciones útiles sin supervisión estricta, construyendo un marco empírico que amplía y refina los fundamentos teóricos originales.

A pesar de la complejidad inherente a la explicación de la evolución de un campo académico dinámico y multidimensional como el aprendizaje automático, resulta útil, a efectos expositivos, identificar fases generales que permiten comprender cómo se ha configurado el camino hacia las tecnologías actuales de aprendizaje profundo y grandes modelos de lenguaje.

3.1. 2000-2010: auge de la ciencia de datos y los modelos especializados

Entre finales de los años noventa y principios de la década de 2010, emergió un paradigma pragmático que hoy se reconoce como Ciencia de Datos. Este enfoque priorizaba la construcción de modelos matemáticos específicos para cada problema, con estructuras relativamente simples, optimización convexa y propiedades estadísticas bien comprendidas. El dominio de esta aproximación se alinea con una mayoría académica que promulga la adherencia a principios teóricos como camino al éxito. Aunque los modelos de aprendizaje automático de la época tenían una precisión limitada debido a su simplicidad matemática —lo que permitía que la teoría explicara su comportamiento y generara expectativas predecibles—, esta misma simplicidad facilitó la creación de modelos predictivos razonablemente eficaces a partir de los conjuntos de datos relativamente pequeños de la época. Esta capacidad fue suficiente para habilitar modelos de negocio dominantes en la actualidad, especialmente aquellos basados en recomendación, personalización y predicción, que sustentaron de manera decisiva el crecimiento y consolidación de empresas como Google, Amazon o Netflix. En este sentido, la aparente limitación impuesta por la adherencia a la teoría de los modelos no impidió su impacto práctico y económico; su predictibilidad y manejabilidad contribuyeron a transformar la industria de la información y sentaron las bases de los sistemas de datos masivos.

Sin embargo, hacia el final de este período, este enfoque comenzó a evidenciar límites claros y estructurales que condicionaron su potencial de innovación. En primer lugar, la necesidad de diseñar modelos matemáticos robustos y *ad hoc* para cada problema específico implicaba costos elevados de desarrollo, tanto en términos de tiempo de investigación como de recursos humanos y computacionales, generando fricciones significativas que restringían la velocidad y amplitud de la innovación. En segundo lugar, se observaron rendimientos decrecientes en términos de precisión: los primeros modelos representaban avances sustanciales respecto a enfoques previos, pero las iteraciones posteriores, aunque refinadas, no lograban incrementos relevantes en la exactitud ni en la utilidad práctica de los sistemas, proporcionando únicamente mejoras marginales que difícilmente justificaban nuevas inversiones significativas. Finalmente, se identificaron límites claros en la escalabilidad: cuando estos modelos cuidadosamente diseñados dentro de los marcos teóricos existentes se aplicaban a dominios de entrada más complejos, como imágenes de alta resolución o grandes volúmenes de texto, su desempeño se degradaba de manera notable y las arquitecturas se rompían, revelando que las soluciones exitosas en problemas controlados no podían generalizar de manera eficiente a escenarios más amplios y

realistas. Estos factores combinados explican por qué el enfoque ad hoc y altamente matemático llega a un estancamiento.

3.2. 2010-2016: *big data*

La transición en la década de 2010 estuvo marcada por un cambio de escala tanto en los datos como en la infraestructura de procesamiento disponible para la investigación y la aplicación del aprendizaje automático. Esta etapa, conocida comúnmente como *big data*, puede entenderse como una evolución natural del paradigma de aprendizaje supervisado basado en la teoría estadística desarrollado entre 2000 y 2010. Durante la fase anterior, los modelos matemáticos específicos para cada problema habían demostrado ser efectivos y relativamente predecibles, permitiendo desarrollar sistemas prácticos de recomendación, predicción y personalización. Sin embargo, estos modelos alcanzaron rápidamente límites estructurales: requerían costos elevados de desarrollo, presentaban rendimientos decrecientes en precisión y problemas de escalabilidad a dominios más complejos, como imágenes de alta resolución o grandes volúmenes de texto.

El fenómeno *big data* surgió como una respuesta técnica y económica a estos límites. Por un lado, la expansión de la captura de datos y la disponibilidad de grandes volúmenes de información permitieron plantear modelos de aprendizaje más ambiciosos, que podrían aprovechar conjuntos de datos mucho mayores que los que habían sido posibles hasta la fecha. Esta expansión no solo incluyó el aumento de la cantidad de datos generados por usuarios, sensores, transacciones comerciales, redes sociales, registros biomédicos y dispositivos conectados, sino también un incremento en la diversidad de aplicaciones.

Para gestionar esta enorme cantidad de datos, se desarrollaron tecnologías específicas de almacenamiento y procesamiento distribuido. Entre las más relevantes destacan los sistemas de archivos distribuidos como Hadoop Distributed File System (HDFS), que permiten almacenar *petabytes* de información de manera redundante y escalable, y *frameworks* de procesamiento distribuido como Hadoop MapReduce y Spark, que facilitan la ejecución de operaciones de transformación y agregación de datos de forma paralela en clústeres de servidores. Estas herramientas, combinadas con el surgimiento de soluciones de almacenamiento en la nube (*cloud computing*) y redes de alta velocidad, crearon un ecosistema capaz de soportar el manejo masivo de información, democratizando el acceso a capacidades que anteriormente solo podían ser asumidas por grandes empresas con infraestructura propia.

Este incremento en la disponibilidad de datos y recursos de procesamiento generó, a su vez, un cambio cultural en la investigación y la industria. La posibilidad de trabajar con conjuntos de datos masivos abrió nuevas oportunidades para el aprendizaje supervisado, pero también evidenció que los enfoques tradicionales basados en aprendizaje estadístico tenían limitaciones para justificar el retorno de la inversión (ROI) de esta

infraestructura. Modelos relativamente simples y *ad hoc*, diseñados para problemas específicos, ya no podían aprovechar plenamente la escala de los datos ni los recursos computacionales disponibles. La industria y la academia comenzaron a percibir que, para aprovechar estos conjuntos de datos y justificar las inversiones en infraestructura, era necesario explorar aproximaciones más generales y empíricas que pudieran capturar patrones complejos sin requerir un diseño ad hoc para cada caso.

De este modo, la fase *big data* constituye un puente crítico hacia el resurgimiento del aprendizaje profundo. Por un lado, consolidó los principios heredados de la teoría de aprendizaje estadístico (Vapnik, 1998; Shalev-Shwartz, 2014): la necesidad de modelos robustos, la relación entre complejidad y tamaño de datos y la importancia de buenas técnicas de optimización como guía teórica. Por otro lado, generó un entorno propicio para experimentación empírica masiva, donde los límites de la teoría se podían observar de manera práctica y reproducible gracias a la abundancia de datos y capacidad de cómputo. Esta combinación de teoría sólida, infraestructura escalable y *datasets* masivos sentó las bases para el desarrollo de arquitecturas neuronales profundas más grandes y complejas, capaces de abordar dominios como visión por computadora y lenguaje natural, que posteriormente conducirían al surgimiento de modelos de lenguaje a gran escala (LLM) y al resurgimiento empírico del aprendizaje profundo.

3.3. 2008-2016: Mafia de la IA y anomalías

Mientras la Ciencia de Datos y el paradigma *big data* se consolidaban en el mercado a partir de técnicas de aprendizaje estadístico bien fundamentadas teóricamente, se desarrollaba en paralelo una línea de investigación menos dominante pero persistente centrada en redes neuronales profundas. Esta corriente, a menudo asociada a un núcleo reducido de investigadores —frecuentemente representados por Geoffrey Hinton, Yann LeCun y Yoshua Bengio, conocidos coloquialmente como la "Mafia de la IA" y galardonados con premios como el Nobel y la Medalla Turing—, aprovechó de manera sistemática el aumento progresivo de la capacidad computacional, el abaratamiento del almacenamiento y la disponibilidad de grandes volúmenes de datos para extender el estudio empírico del aprendizaje profundo (LeCun *et al.*, 2015; Bengio *et al.*, 2013). A diferencia del enfoque estadístico clásico, estas investigaciones exploraron la combinación de técnicas supervisadas y no supervisadas con el objetivo de aprender representaciones internas jerárquicas y reutilizables, capaces de capturar regularidades latentes en los datos. El resultado fundamental de estos esfuerzos fue la constatación empírica de que los modelos de aprendizaje profundo podían superar de forma consistente la precisión de los enfoques estadísticos tradicionales en dominios donde estos últimos habían alcanzado un estancamiento prolongado, como el reconocimiento de imágenes, el procesamiento del habla y, posteriormente, el lenguaje natural (Hinton *et al.*, 2006). Este avance no supuso una negación explícita de los principios del aprendizaje estadístico, sino una extensión práctica de sus límites observables, demostrando que bajo condiciones adecuadas de

datos y cómputo era posible aprender representaciones robustas que transformaban problemas previamente intratables en tareas abordables mediante modelos entrenables de extremo a extremo.

Cabe destacar que, durante este período, comenzaron a acumularse una serie de observaciones empíricas clave que resultaban difíciles de ignorar desde el marco teórico dominante en ese momento. Estos resultados, reproducidos de manera consistente en distintos laboratorios y dominios de aplicación, se manifestaban como anomalías empíricas que no podían explicarse adecuadamente a partir de la teoría del aprendizaje estadístico clásica ni de sus supuestos sobre complejidad, generalización y escalabilidad. Lejos de tratarse de excepciones aisladas, estas observaciones señalaron patrones sistemáticos en el comportamiento de los modelos neuronales profundos bajo regímenes de datos y cómputo masivos, dando lugar a una nueva tendencia de investigación centrada en comprender y explotar estos fenómenos. En las secciones siguientes se detallan dichas observaciones empíricas y se analiza su papel en la transformación del campo del aprendizaje automático.

3.3.1. Aprendizaje de representaciones mediante observación no supervisada

Una de las primeras observaciones empíricas que contribuyó de manera decisiva al resurgimiento del aprendizaje profundo fue la constatación de que es posible aprender estructuras internas útiles para la resolución de problemas complejos mediante técnicas no supervisadas, es decir, a partir de la mera observación de los datos sin necesidad de anotaciones humanas explícitas. Este hallazgo contradecía una de las limitaciones prácticas más relevantes de los enfoques dominantes hasta ese momento: la fuerte dependencia de grandes volúmenes de datos etiquetados, cuyo coste de obtención constituía un cuello de botella tanto en investigación como en aplicaciones industriales.

El trabajo seminal de Hinton y Salakhutdinov (Hinton, 2006) demostró que redes neuronales profundas, entrenadas de manera no supervisada para reducir la dimensionalidad de los datos, eran capaces de aprender representaciones latentes que capturaban regularidades estructurales relevantes del dominio subyacente. Estas representaciones no solo preservaban información esencial, sino que facilitaban de forma significativa el posterior aprendizaje supervisado, permitiendo alcanzar mejores resultados con una fracción de los ejemplos anotados previamente necesarios.

Desde una perspectiva sistémica, este enfoque introdujo un cambio cualitativo en la manera de diseñar algoritmos de aprendizaje. En lugar de construir modelos específicos para cada tarea desde cero, se hizo posible reutilizar modelos preentrenados como componentes genéricos de representación, adaptables a múltiples problemas mediante un ajuste fino posterior. Esta reutilización redujo de manera drástica el esfuerzo de ingeniería algorítmica, disminuyó la dependencia de datos etiquetados y sentó las bases concep-

tuales de los actuales modelos fundacionales, marcando un punto de inflexión respecto a los paradigmas de aprendizaje supervisado especializados que habían dominado la disciplina hasta entonces.

3.3.2. Escalado de datos y arquitecturas profundas en visión artificial

Una segunda observación empírica fundamental emergió a partir de la convergencia entre la disponibilidad de conjuntos de datos de una escala sin precedentes y la aplicación de arquitecturas neuronales profundas entrenadas de manera eficiente. El proyecto ImageNet constituyó, en este sentido, la mayor base de datos de imágenes naturales creada hasta la fecha con fines de investigación, proporcionando millones de ejemplos etiquetados que cubrían una amplia diversidad de objetos y contextos visuales. Este recurso transformó el estudio de la visión artificial al situar el problema en un régimen de datos radicalmente distinto al previamente explorado.

Sobre esta base empírica, el trabajo de Krizhevsky *et al.* (2012) demostró que una arquitectura neuronal conocida —las redes convolucionales profundas, previamente exploradas sin éxito concluyente en décadas anteriores—, combinada con innovaciones clave en los procedimientos de entrenamiento y el uso intensivo de *hardware* especializado, podía superar de forma contundente a todos los algoritmos de visión artificial desarrollados en los treinta años previos. El modelo conocido como AlexNet redujo de manera drástica el error en la competición ImageNet, marcando una ruptura clara con el estado del arte dominado hasta entonces por métodos basados en características diseñadas manualmente y clasificadores estadísticos especializados.

Este resultado tuvo un impacto profundo en la comunidad científica al evidenciar que los límites observados en los enfoques estadísticos clásicos no respondían necesariamente a barreras fundamentales del problema, sino a restricciones impuestas por la escala de los datos, la capacidad de los modelos y los métodos de optimización disponibles. La combinación de grandes volúmenes de datos, arquitecturas profundas y capacidad de cómputo suficiente reveló un nuevo régimen empírico en el que la precisión y la capacidad de generalización continuaban mejorando de forma sistemática, desafiando las expectativas establecidas por la teoría dominante y consolidando al aprendizaje profundo como la aproximación central en visión por computador.

3.3.3. Arquitecturas profundas como generadores de contenido

Una de las contribuciones más significativas del aprendizaje profundo en este tiempo fue demostrar que los modelos neuronales no solo podían aprender representaciones útiles para clasificación o predicción, sino también generar contenido de manera autónoma y creativa, guiada por objetivos definidos por el usuario. Este avance se materializó con las

redes generativas antagónicas (GANs), introducidas por Goodfellow *et al.* (2014), que combinan un generador y un discriminador en un esquema de entrenamiento competitivo. Mediante este mecanismo, los modelos pueden producir ejemplos sintéticos que imitan las propiedades estadísticas del conjunto de datos de entrenamiento, pero además permiten al usuario dirigir la generación hacia características deseadas, ya sea en imágenes, audio u otros tipos de datos.

El impacto de este hallazgo es doble. Por un lado, confirma que las redes profundas aprenden representaciones internas suficientemente ricas como para producir contenido coherente y novedoso, incluso en dominios donde los métodos estadísticos clásicos no podían generar resultados útiles sin supervisión intensiva. Por otro lado, abre un nuevo paradigma de aplicaciones prácticas: la generación creativa dirigida por el usuario, que incluye desde síntesis de imágenes artísticas hasta la creación de datos sintéticos para entrenamiento y validación de otros modelos, o interfaces interactivas donde el aprendizaje profundo responde a instrucciones humanas. En consecuencia, esta observación ilustra cómo el aprendizaje profundo no solo amplía los límites de precisión y generalización, sino que también redefine lo que significa "aprender" y "crear" dentro de un marco computacional, consolidando una nueva categoría de anomalías empíricas que desafían la teoría estadística tradicional.

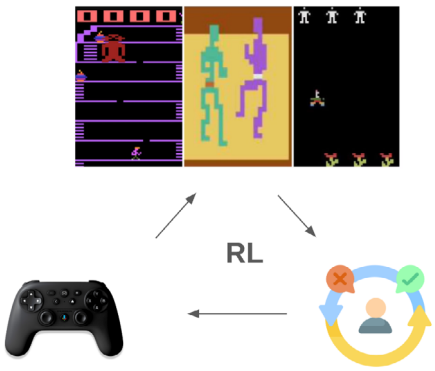
3.3.4. Redes neuronales profundas, aprendizaje por refuerzo y simuladores

Una observación empírica crítica durante la consolidación del aprendizaje profundo fue la demostración de que, dado un sistema de control general, un simulador suficientemente realista del sistema a controlar y un criterio de retroalimentación adecuado, es posible entrenar modelos de aprendizaje por refuerzo profundo para generar soluciones de control efectivas sin intervención humana directa. El trabajo seminal de Mnih *et al.* (2015) ilustró este principio mediante la aplicación de redes neuronales profundas a entornos de videojuegos complejos, mostrando que, con entrenamiento suficiente en episodios simulados y utilizando una señal de recompensa bien diseñada, los agentes podían aprender políticas de control que superaban el rendimiento humano en varias tareas.

El elemento clave de este avance radica en la combinación de tres factores: primero, la disponibilidad de un simulador que reproduzca de manera realista la dinámica del sistema, permitiendo la acumulación de experiencia en millones de episodios sin riesgos físicos; segundo, la formulación de un criterio de retroalimentación que guíe el aprendizaje hacia comportamientos deseables; y tercero, la capacidad de la red neuronal profunda para construir representaciones internas ricas y jerárquicas del problema, que facilitan la planificación de acciones complejas y creativas. Esta arquitectura permitió que los modelos no solo aprendieran soluciones de control óptimas, sino que también desarrollaran secuencias de acciones novedosas y efectivas, aplicables en entornos del mundo real.

FIGURA 2

EL APRENDIZAJE POR REFUERZO COMBINADO CON SIMULADORES Y UNA FUNCIÓN OBJETIVO BIEN DEFINIDA CONSTITUYEN EL MARCO DE SISTEMAS DE CONTROL MODERNOS



Fuente: Volodymyr *et al.* (2015).

En conjunto, este hallazgo evidencia que los sistemas de aprendizaje profundo, cuando se combinan con simulación de alta fidelidad y señales de retroalimentación adecuadas, pueden transformar problemas de control general en tareas abordables mediante aprendizaje autónomo. Esto abre la puerta a aplicaciones prácticas en robótica, vehículos autónomos y sistemas de automatización industrial, demostrando que el aprendizaje por refuerzo profundo no solo extiende las capacidades predictivas de las redes neuronales, sino que también permite generar comportamiento creativo y adaptativo en dominios complejos y dinámicos.

3.3.5. Ciencia de Datos 2.0

El período comprendido entre 2016 y 2025 puede caracterizarse como una fase emergente de la disciplina que hemos denominado Ciencia de Datos 2.0. Esta etapa se fundamenta en la convergencia de anomalías empíricas observadas en los años anteriores y la disponibilidad de recursos tecnológicos sin precedentes. Las observaciones previas sobre la capacidad de los modelos neuronales profundos para aprender representaciones útiles, generar contenido creativo y controlar sistemas complejos mediante aprendizaje por refuerzo, dieron lugar a una nueva tendencia: diseñar arquitecturas neuronales profundas robustas, incrementar su tamaño y complejidad, y combinarlas con conjuntos de datos masivos y heterogéneos. La práctica habitual de preentrenamiento no supervisado, seguida de optimización supervisada o reforzada, permite aprovechar patrones subyacentes en los datos y facilita la generalización, reproduciendo de manera consistente fenómenos que antes eran considerados anomalías (Doshi *et al.*, 2019; Hooker *et al.*, 2019).

Entre estas anomalías destaca el fenómeno conocido como *grokking*, en el que modelos suficientemente complejos, expuestos a grandes cantidades de datos y entrenados durante tiempos prolongados con esquemas de regularización adecuados, aprenden de manera estructural la solución subyacente a un problema, separando claramente memorizar de generalizar. La reproducibilidad de este tipo de eventos permite avanzar en la comprensión teórica de los límites del aprendizaje profundo y extender la teoría estadística clásica, integrando observaciones empíricas que previamente se consideraban excepciones (Doshi *et al.,* 2019).

El éxito de la Ciencia de Datos 2.0 no puede separarse de la coincidencia con avances en *hardware*, fenómeno denominado *hardware lottery* (Hooker, 2021). La disponibilidad de GPUs de alta eficiencia, almacenamiento masivo y arquitectura distribuida permitió que estas redes complejas pudieran entrenarse en tiempos razonables, haciendo visibles los efectos observables y reproducibles que definieron la tendencia. Es importante notar que los modelos actuales no constituyen la única vía posible para desarrollar tecnologías de este tipo; son, más bien, aquellos que pueden implementarse eficazmente con el *hardware* existente, condicionando la evolución empírica y la adopción industrial.

El ejemplo más público de Ciencia de Datos 2.0 es ChatGPT, donde un protocolo de aprendizaje por refuerzo orientado a "jugar el juego del lenguaje" fue expuesto al mundo, acumulando millones de episodios lingüísticos y permitiendo que un modelo de lenguaje a gran escala aprenda una política efectiva de interacción con el mundo (Long *et al.*, 2022). En este caso, los principios algorítmicos originales aplicados a videojuegos, como los agentes de Atari, se mantuvieron intactos; únicamente el controlador del juego fue sustituido por tokens de lenguaje, y el juego consiste en interaccionar de manera coherente y productiva en un entorno lingüístico abierto. Este enfoque demuestra que la combinación de arquitecturas profundas, grandes conjuntos de datos de entrenamiento y algoritmo de retropropagación del error pueden producir sistemas capaces de comportamientos complejos, creativos y útiles, consolidando Ciencia de Datos 2.0 como la etapa contemporánea dominante en inteligencia artificial.

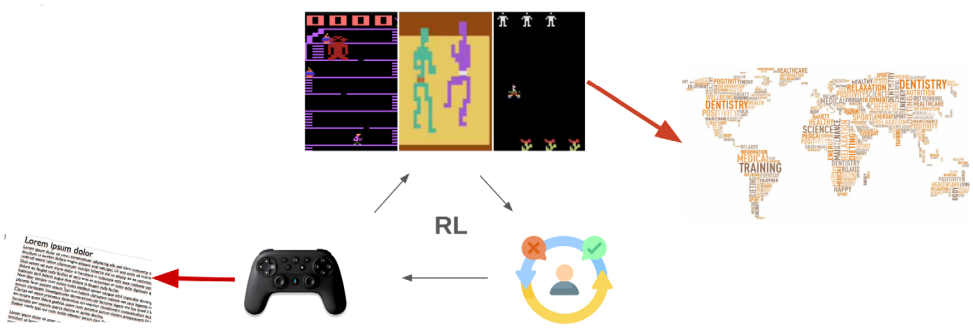
4. PANARQUÍA ARTIFICIAL

Tras la revisión de los últimos veinticinco años de avances en inteligencia artificial y aprendizaje estadístico, resulta evidente que lo que a primera vista puede interpretarse como un fracaso de una tecnología y un éxito de otra constituye, en realidad, un conjunto de avances incrementales que son teóricamente coherentes y complementarios en la práctica. La teoría y la experiencia acumuladas muestran que existen numerosos problemas que aún se resuelven de manera más eficiente mediante enfoques de aprendizaje estadístico, lo que explica que la industria asociada a estos métodos continúe activa y próspera.

Visualizar el panorama de la IA en términos de tecnologías ganadoras o perdedoras, como propone la narrativa de "veranos e inviernos", conduce con frecuencia a decisiones equivocadas en múltiples niveles, desde la inversión hasta la formulación de políticas públicas. Si bien es cierto que, como hemos detallado, la aparición de resultados empíricos significativos puede generar cambios de perspectiva y reorganización en el campo, esto no implica que los avances previos hayan perdido completamente su valor ni que los nuevos desarrollos reemplacen de manera inmediata las dinámicas existentes. Por ello, para evitar juicios sesgados derivados de un enfoque estrictamente estacional, resulta necesario adoptar un marco de análisis que considere la interacción entre sociedad, tecnología y acontecimientos científicos. Dicho marco permitiría evaluar de manera integral el estado de las tecnologías disponibles en cada momento, reconociendo simultáneamente la continuidad de los conocimientos previos, la emergencia de innovaciones y la influencia de factores socioeconómicos en la adopción y desarrollo de estas. Una aproximación de este tipo posibilitaría decisiones más informadas, equilibradas y estratégicas en la planificación de investigación, inversión y políticas públicas en el ámbito de la inteligencia artificial.

FIGURA 3

LOS MODELOS DE LENGUAJE ACTUALES PROVIENEN DE UN USO EFECTIVO DE TÉCNICAS DE APRENDIZAJE POR REFUERZO EN EL CASO DEL LENGUAJE NATURAL



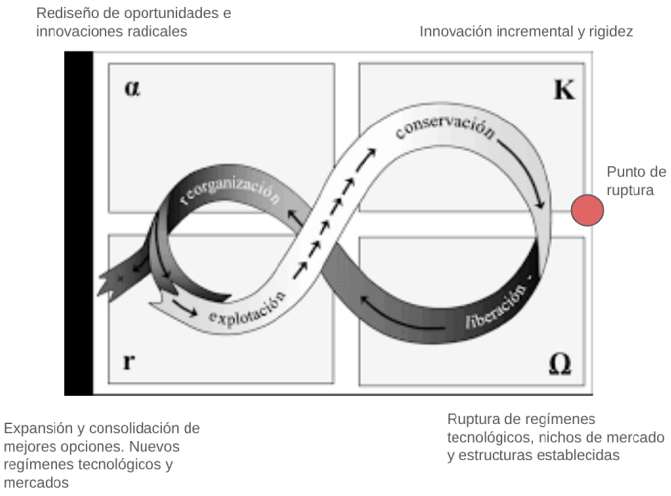
5. PANARQUÍA Y CICLOS ADAPTATIVOS EN INTELIGENCIA ARTIFICIAL

El concepto de *panarquía* y los *ciclos adaptativos* provienen de la teoría de sistemas complejos y fueron desarrollados inicialmente en el contexto de la ecología por Holling y colaboradores (Holling, 2001). Su objetivo principal es describir cómo los sistemas complejos, ya sean naturales, sociales o tecnológicos, evolucionan a través de fases recurrentes de estabilidad, colapso y reorganización. La intuición básica es que lo que puede parecer un comportamiento caótico o impredecible a nivel de un sistema específico puede, de hecho, estar estructurado por ciclos de adaptación que operan en distintas escalas, interactuando entre sí.

En general, un *ciclo adaptativo* puede entenderse como un proceso de cuatro fases: *conservación*, *liberación*, *reorganización* y *explotación*. Estas fases evolucionan de manera continua e indefinida, tal y como representa la figura 4. En la fase de *conservación*, los recursos, conocimientos y estructuras del sistema se consolidan, aumentando su eficiencia y especialización, pero al mismo tiempo incrementando su rigidez frente a perturbaciones externas. Posteriormente, la fase de *liberación* ocurre cuando una perturbación significativa, interna o externa, rompe la estabilidad del sistema, liberando los recursos acumulados y permitiendo que las restricciones anteriores desaparezcan. Esta liberación da paso a la fase de *reorganización*, en la que los elementos del sistema se reconfiguran, se prueban nuevas interacciones y emergen nuevas estrategias de funcionamiento. Finalmente, durante la fase de *explotación*, las innovaciones y reconfiguraciones exitosas se consolidan, los recursos se utilizan de manera eficiente, y el sistema entra nuevamente en un período de conservación, cerrando el ciclo. Cada fase es esencial: la conservación de recursos y conocimiento permite estabilidad y eficiencia, mientras que la liberación y reorganización facilitan la innovación y la adaptabilidad.

FIGURA 4

VISUALIZACIÓN DE UN CICLO ADAPTATIVO



El modelo de ciclos adaptativos ha sido utilizado con éxito para explicar fenómenos complejos en contextos muy diversos, desde ecosistemas forestales y pesqueros (Holling, 1995), hasta sistemas socioeconómicos como mercados financieros y organizaciones urbanas (Gunderson and Holling, 2002). En todos estos casos, los cambios que a primera vista parecen abruptos o desordenados pueden interpretarse de manera más precisa como transiciones entre ciclos recurrentes de conservación y reorganización.

Aplicando esta perspectiva al campo de la inteligencia artificial, podemos reinterpretar la narrativa de los "veranos e inviernos" como manifestaciones superficiales de ciclos adaptativos.

Durante reorientaciones fundamentales en la IA como campo de conocimiento, como los finales de los años 80 con los modelos neuronales iniciales o los recientes resurgimientos del aprendizaje profundo y los modelos de lenguaje a gran escala, el sistema científico pasa por fases de reorganización y explotación iniciadas por una experiencia que rompe con *statu quo* científico (conservación) anterior.

Las invenciones generadas se estudian de manera intensiva, los recursos (datos, computación y capital humano) se concentran y la eficiencia de los sistemas recientemente ideados aumenta. Sin embargo, estas fases también van generando rigidez: limitaciones teóricas, cuellos de botella computacionales y dificultades de generalización, etc. que eventualmente conducen a una nueva fase de liberación. Durante esta nueva fase de reorganización el ciclo no se frena sino que se repite. Investigadores y empresas exploran nuevas arquitecturas, combinan técnicas supervisadas y no supervisadas, desarrollan infraestructuras de *Big Data* o experimentan con simuladores y aprendizaje por refuerzo para redefinir lo que es posible. Así emergen los avances que posteriormente entrarán en explotación: redes neuronales profundas escaladas, modelos generativos como GANs, y más recientemente, modelos de lenguaje a gran escala que interactúan con el mundo a través de retroalimentación humana (como son ChatGPT y otros sistemas comerciales similares). Es importante destacar que cada ciclo no elimina ni rompe completamente el valor de los desarrollos anteriores. Por ejemplo, los métodos estadísticos y los modelos especializados continúan siendo hoy útiles en dominios donde su eficiencia e interpretabilidad son superiores y comparten raíces teóricas con los modelos neuronales.

Si consideramos el avance en IA como una serie de ciclos adaptativos y aspiramos a tener un modelo completo, no podemos ignorar la percepción de "veranos e inviernos", sino que necesitamos ampliar el modelo de manera que esta percepción sea explicable. Para ello es necesario reconocer y modelar que los ciclos de invención en IA no ocurren en un sistema aislado, sino que influencian y son influenciados por otros ciclos adaptativos con escalas espaciales y temporales que pueden ser diferentes. Este es el rol del concepto de *panarquía*, el cual añade una dimensión adicional al permitir observar múltiples ciclos adaptativos interactuando a distintas escalas. El observar múltiples ciclos adaptativos de manera conectada ayuda a reconciliar situaciones paradójicas como narrativas de "invierno" acerca de tecnologías que están avanzando o "verano" acerca de tecnologías cercanas a su límite de rendimiento. En la siguiente sección se introduce nuestro modelo de panarquía para el caso de la IA.

TABLA 1

PRINCIPALES FASES EN EL RESURGIMIENTO DEL APRENDIZAJE PROFUNDO

	Fase	**Ciencia de Datos**
2000-2010	Observaciones científicas	Modelos especializados simples, enfoque *ad hoc*, optimización convexa, propiedades estadísticas bien comprendidas
	Impacto práctico	Sistemas de recomendación y personalización; consolidación de empresas como Google, Amazon y Netflix
	Herramientas/Técnicas	Modelos lineales y bayesianos, *datasets* pequeños
	Fase	***Big Data***
2010-2016	Observaciones científicas	Disponibilidad masiva de datos y cómputo; límites estructurales de modelos estadísticos evidenciados
	Impacto práctico	Infraestructura distribuida y democratización del procesamiento masivo; puente hacia arquitecturas neuronales profundas
	Herramientas/Técnicas	Hadoop, Spark, HDFS, *cloud computing*; aprendizaje supervisado escalable
	Fase	**Mafia de la IA y anomalías**
2008-2016	Observaciones científicas	Redes neuronales profundas aprenden representaciones jerárquicas; anomalías reproducibles; superan modelos estadísticos en visión, lenguaje y habla
	Impacto práctico	Éxitos en visión por computador (ImageNet), generación de contenido (GANs), aprendizaje por refuerzo; reducción de dependencia de datos etiquetados
	Herramientas/Técnicas	Redes convolucionales profundas, preentrenamiento no supervisado, GANs, simuladores para RL
	Fase	**Ciencia de Datos 2.0**
2016-2025	Observaciones científicas	Consolidación de anomalías empíricas; preentrenamiento seguido de ajuste fino; fenómenos como *grokking*
	Impacto práctico	Modelos fundacionales escalables (LLMs, ChatGPT); aplicaciones en lenguaje, control y generación de contenido
	Herramientas/Técnicas	Arquitecturas profundas masivas, GPUs, cómputo distribuido, aprendizaje por refuerzo en lenguaje natural

5.1. Panarquía artificial

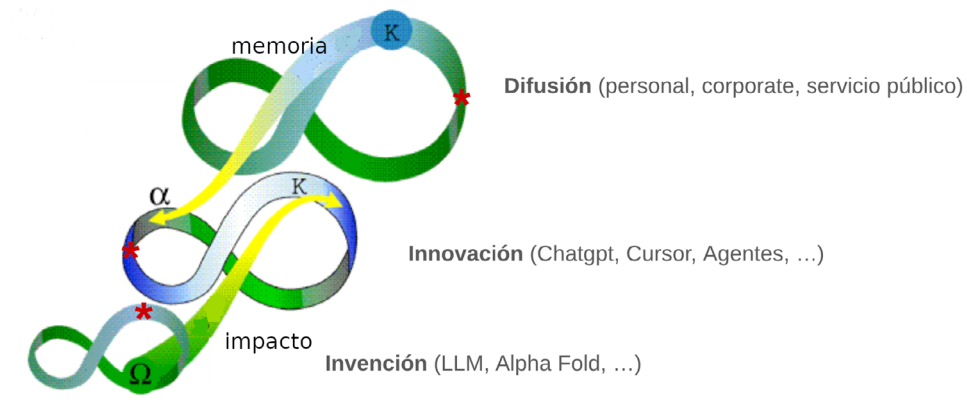
La entrada de los modelos de lenguaje a gran escala (LLM) en la esfera pública ha generado una disonancia notable entre expectativas, regulación y realidad. Por un lado, se han propagado afirmaciones sobre la cercanía de una superinteligencia de uso general

(*Artificial General Intelligence* o AGI) con propuestas de regulaciones orientadas a la no proliferación similares al caso de la tecnología nuclear bélica. Además, se han movilizado niveles de inversión sin precedentes para financiar su desarrollo y despliegue. Sin embargo, la adopción de la tecnología, varios años después de su irrupción inicial, no ha generado los cambios radicales esperados; los retornos sobre el capital invertido siguen siendo inciertos, y aunque su uso es amplio, se limita a aplicaciones muy específicas. La coexistencia de estas realidades ha generado una polarización hacia dos extremos, uno de confianza ciega en la proximidad de una AGI y otro de espera por el colapso inevitable de la tecnología.

El análisis de esta situación compleja sin recurrir a percepciones requiere un marco analítico más completo, capaz de reconciliar las dinámicas de invención, innovación y consolidación tecnológica. Proponemos interpretar la situación a través de un modelo de *panarquía multinivel*, compuesto por ciclos adaptativos interconectados a diferentes escalas, tal y como se representa en la figura 5. La división en niveles propuesta se inspira en la clasificación sugerida por Narayanan y Kapoor (2025):

FIGURA 5

PANARQUÍA ARTIFICIAL Y ESTADO A FECHA DE PUBLICACIÓN DE ESTE LIBRO



1. *Ciclo de difusión*: abarca la adopción de la tecnología en la vida cotidiana, en diversos ámbitos como el personal, el corporativo y los servicios públicos. Este nivel refleja la integración progresiva de la IA en rutinas establecidas, regulaciones y estructuras sociales existentes, donde la inercia humana y los marcos regulatorios actúan como filtros que ralentizan la adopción, pero también protegen contra riesgos potenciales.

2. *Ciclo de innovación:* corresponde a la transformación de la tecnología en productos utilizables y servicios concretos, como ChatGPT, Cursor u otras aplicaciones comerciales basadas en LLM. Aunque los ciclos de innovación pueden ser más ágiles que los de difusión, es importante tener en cuenta que la implementación de productos innovadores conlleva tiempos e inercias, incluso cuando se utilizan tecnologías base plenamente funcionales.

3. *Ciclo de invención*: refleja los ciclos generados por descubrimientos en laboratorios de investigación o empresas específicas, donde surgen nuevas tecnologías fundacionales, como los propios LLM.

Desde esta perspectiva multinivel, resulta más sencillo comprender la situación actual. A nivel de invención, los ciclos recientes han reforzado la supremacía del aprendizaje profundo, limitando temporalmente la exploración de caminos alternativos. Los avances son incrementales y se concentran en perfeccionar las líneas de investigación dominantes, lo que genera rigidez tecnológica y focaliza los recursos —tanto de capital como humanos— en la expansión y consolidación de esta tecnología.

A nivel de innovación, la tecnología se ha encontrado con limitaciones: a) para trasladar sus capacidades a aplicaciones prácticas concretas y b) para alcanzar niveles de rendimiento suficientes que permitan su uso en problemas de mayor impacto, como son las tecnologías de agentes. Aunque el avance hacia casos de uso de mayor impacto ha sido más lento que el ciclo de invención, el ciclo de innovación está superando su fase de reorganización. La mayoría de productos digitales de consumo han reevaluado su posicionamiento a la luz de las nuevas capacidades y se vislumbra una fase de explotación, en la que la tecnología puede alcanzar una penetración masiva y consolidar las mejores prácticas de diseño y despliegue.

Finalmente, en el ciclo de difusión, la adopción está modulada por regulaciones, inercia institucional y estructuras sociales consolidadas, que retrasan la penetración completa de la tecnología, pero al mismo tiempo actúan como un mecanismo de mitigación de riesgos.

Mientras la narrativa de "veranos e inviernos" promueve un enfoque reactivo, donde las decisiones se toman en función de percepciones de éxito o fracaso a corto plazo y en una sola escala temporal, la perspectiva de *panarquía* permite un manejo más estratégico de la evolución tecnológica. Al analizar los distintos niveles de invención, innovación y difusión en relación con la fase del ciclo adaptativo en la que se encuentran, es posible diseñar intervenciones y estrategias mejor contextualizadas. Por ejemplo, las regulaciones existentes, las prácticas institucionales consolidadas y la experiencia adquirida durante difusiones tecnológicas previas ya actúan como mecanismos de memoria (*remember*) que orientan la implantación en IA, evitando riesgos innecesarios y promoviendo desarrollos más sostenibles. De manera complementaria, una vez que la fase de conservación actual en la investigación en aprendizaje profundo llegue a su límite, se espera que nuevas observaciones disruptivas den lugar a nuevos ciclos de investigación,

utilizando los recursos, conocimiento y aprendizajes acumulados. La visión de panarquía no solo explica la coexistencia de estabilidad social, innovación paulatina e invención disruptiva en el ecosistema de la inteligencia artificial, sino que también proporciona un marco operativo para gestionar su evolución de manera proactiva, reconociendo la interacción de diferentes ciclos a diferentes escalas.

5.1.1. Panarquía artificial actual

Una manera de poner a prueba el marco de panarquía consiste en analizar de manera prospectiva la situación actual de la inteligencia artificial, considerando la interacción de ciclos adaptativos a distintos niveles.

En el nivel de invención, es de esperar que la fase de conservación actual se vea interrumpida por algún avance súbito o por un límite significativo. Entre los factores que podrían provocar esta ruptura se incluyen:

- Modelos pequeños y manejables que progresivamente demuestran capacidades cercanas a modelos propietarios de gran escala, cuestionando la necesidad del consumo de grandes recursos de computación.

- Dificultades para encontrar fuentes de energía o infraestructura de *hardware* que soporten modelos más grandes, generando cuellos de botella que podrían ralentizar la evolución.

Paralelamente, la investigación en esta fase se ha basado en *benchmarks* académicos que en un primer momento, parecían resolver problemas prácticos deseados. Sin embargo, pronto se observó que los avances académicos no se traducían en productos útiles, lo que motivó el desarrollo de *benchmarks* más realistas y orientados a aplicaciones prácticas. Este desfase entre los objetivos de la investigación académica y las expectativas de innovación explica en gran parte el avance lento en la difusión de la tecnología percibido recientemente. Esta brecha ha comenzado a cerrarse progresivamente, aunque todavía persisten limitaciones significativas en muchos casos de uso deseados.

En el ciclo de innovación, se ha superado la fase de reorganización, en la que se han detectado limitaciones concretas de la tecnología a la hora de crear aplicaciones prácticas. Esta fase de experimentación, aunque puede interpretarse de manera pesimista, ha generado guías de diseño más robustas para sistemas prácticos basados en IA, preparando el terreno para una fase de explotación. Entre los ejemplos recientes de limitaciones catalogadas destacan:

- Limitaciones en la capacidad de razonamiento frente a problemas complejos (Shojaee *et al.,* 2025).

- Dificultades en sistemas de frontera para mantener consistencia y coherencia en tareas de razonamiento general (Kamradt *et al.,* 2025).

- Pérdida de coherencia en conversaciones multiturno de modelos de lenguaje (Zhou *et al*., 2025).

- Restricciones en el manejo de contextos extensos, limitando la aplicabilidad práctica de LLMs en tareas complejas (Kriman *et al.*, 2025).

Por otra parte, las expectativas financieras y valoraciones bursátiles de la innovación en IA pueden actuar como disruptor negativo en el ciclo de innovación actual. Una corrección significativa de expectativas de retorno de inversión podría conducir a un abandono masivo de la tecnología.

Finalmente, en el ciclo de difusión, existen limitaciones naturales que han retrasado la adopción masiva de la IA en la vida pública:

- No todo el conocimiento humano está formalizado o escrito; gran parte es tácito y social, dificultando su uso por sistemas de IA.

- Rigidez en organizaciones, procesos de decisión y estructuras institucionales que retrasan la integración de nuevas tecnologías.

- Automatización parcial: solo alrededor del 10 % de muchos procesos puede ser efectivamente automatizado con IA.

- Problemas de gestión del cambio y fricciones regulatorias, tanto sociales como externas (por ejemplo, limitaciones energéticas).

Estas barreras explican por qué la disrupción social esperada aún no ha ocurrido de manera masiva. Sin embargo, desde la perspectiva de los ciclos adaptativos, es previsible que la tecnología eventualmente se incorporará de manera más profunda en la sociedad. La interacción entre ciclos de invención, innovación y difusión indica que, aunque la adopción plena está siendo gradual, el avance se está produciendo en todos los niveles de la panarquía planteada.

6. CONCLUSIONES

El análisis de la inteligencia artificial mediante el marco de panarquía y ciclos adaptativos nos permite comprender de manera más matizada la dinámica tecnológica actual. La narrativa de "veranos e inviernos" es insuficiente para capturar la complejidad de la interacción entre investigación, innovación y adopción social.

Los avances disruptivos son resultado de una acumulación de observaciones de base científica que discuten una teoría dominante hasta reorientar nuestra comprensión de un campo de conocimiento e iniciar una reorganización de su estudio; el ciclo científico no avanza de manera aislada sino que interactúa con ciclos de innovación que generan productos con alto impacto en la sociedad; finalmente los ciclos de difusión, evidenciados a través de barreras institucionales, bloqueos comerciales e inercias sociales modulan la penetración de una tecnología, con consecuencias tanto positivas como negativas. Todo análisis o regulación novedosa en ámbitos sin precedentes —como la GDPR en el caso de la expansión del intercambio de datos o el AI Act en el caso de la IA—incorpora, de forma explícita o implícita, un modelo de evolución futura. Cuando dicho modelo no refleja de manera realista y exhaustiva el fenómeno que se pretende regular, existe el riesgo de desaprovechar oportunidades o de generar dinámicas no deseadas.

La perspectiva multinivel presentada ofrece una base sólida para analizar y anticipar la evolución de la IA a corto y medio plazo, integrar aprendizajes previos y gestionar riesgos y oportunidades de manera estratégica y contextualizada.

Referencias

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8),1798–1828.

Doshi,D., Das A., He, T., and Gromov, A. (2019). To grok or not to grok: Disentangling generalization and memorization on corrupted algorithmic datasets. *arXiv preprint arXiv:1912.*12349.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial networks. In *Advances in Neural Information Processing Systems*, volume 27.

Gunderson, L. H., and Holling, C. S. (2002). *Panarchy: Understanding Transformations in Human and Natural Systems*. Island Press.

Hinton, G. E., Osindero, S., and Yee-Whye Teh, Y-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7),1527–1554.

Holling, C. S. (1995). What barriers? what bridges? *Ecological Applications*, *5*(2), 523–525.

Holling, C. S. (2001). Understanding the complexity of economic, ecological, and social systems. *Ecosystems*.

Hooker, S. (2021). The hardware lottery. *Communications of the ACM, 64*(12), 56–63.

Hooker, S., Courville, A., Clark, G., and Dauphin, Y. (2019). What do compressed deep neural networks forget? *arXiv preprint arXiv:1910.* 01742.

Kamradt, G., Landers, B., Pinkard, H., Chollet, F., and Knoop, M. (2025). Arc-agi-2: A new challenge for frontier ai reasoning systems. *arXiv preprint arXiv:2501.00003*.

Kriman, S., Acharya, S., Hsieh, C. P., Sun, S. (2025). Ruler: What's the real context size of your long-context language models? *arXiv preprint arXiv:2501.*00005.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 25, 1097–1105.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation, 1*(4), 541–551.

Meadows, D. H., and Wright, D. (2008). *Thinking in Systems: A Primer*. Chelsea Green Publishing.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Mnih, V. *et al*. (2015). Human-level control through deep reinforcement learning. *Nature*.

Narayanan, A., and Kapoor, S. (2025). *Ai as normal technology: An alternative to the vision of AI as a potential superintelligence*. Knight First Amendment Institute.

Ouyang, L. *et al.* (2022). Training language models to follow instructions with human feedback. *arXiv*.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.

Shalev-Shwartz, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.

Shojaee, P. *et al.* (2025). The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity. *arXiv preprint arXiv:2501.*00002.

Vapnik, V. (1998). *Statistical Learning Theory*. Wiley.

Wolpert, D., and Macready, W. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation.*

Zhou, Y., Neville, J., Laban, P., and Hayashi, H. (2025). Llms get lost in multi-turn conversation. *arXiv preprint arXiv:2501.*00004.