

Módulo 1. Fundamentos e historia de la inteligencia artificial




¿Qué es la inteligencia artificial?


La inteligencia artificial (IA) se define como una rama de la informática que busca desarrollar sistemas capaces de imitar la inteligencia humana. En esencia, implica la creación de algoritmos y programas que puedan aprender, razonar y resolver problemas de forma análoga a un ser humano (Tableau, 2023). Lejos de la visión de ciencia ficción del siglo XX, la IA se ha integrado en nuestra vida cotidiana; sus aplicaciones se han expandido a tal punto que prácticamente todas las facetas de la vida moderna se ven beneficiadas por ella (IA Canarias, 2025).


Muchos interactuamos con tecnologías de IA a diario sin ser plenamente conscientes de ello. Los asistentes virtuales en los teléfonos móviles, como Siri o Google Assistant, que responden a comandos de voz; los sistemas de navegación GPS que calculan rutas óptimas en tiempo real; los filtros de correo no deseado (*spam*) que protegen nuestras bandejas de entrada; y los sistemas de recomendación que sugieren películas en Netflix o productos en Amazon son todos ejemplos palpables de la IA en acción. De hecho, estudios recientes indican que cerca del 80% de las personas ya


utilizan alguna forma de IA en sus actividades diarias (González, 2024). Esta omnipresencia subraya la importancia de comprender no solo qué es la IA, sino también cómo funciona, cuáles son sus fundamentos y cómo ha evolucionado hasta convertirse en la tecnología definitoria de nuestra era.

 1. La IA como disciplina

 2. ¿Cuáles son los principales fundamentos?

 3. Breve historia de la IA: hitos y evolución

 4. Análisis de la actualidad

 Referencias

 Descarga en PDF

1. La IA como disciplina

Para abordar el estudio de la inteligencia artificial, es importante realizar una distinción que a menudo se pasa por alto en la conversación popular: la diferencia entre la IA como un amplio campo de estudio y los «modelos de IA» como aplicaciones específicas y funcionales.

La inteligencia artificial no es un objeto, un robot ni un programa único. Es una disciplina científica y de ingeniería de gran alcance, comparable a campos como la biología, la física o la medicina. Como tal, abarca un ecosistema diverso de subdisciplinas, cada una con objetivos y metodologías propias. Entre estas áreas se encuentran el aprendizaje automático (machine learning), el procesamiento del lenguaje natural (NLP), la visión por computadora, la robótica y los sistemas expertos. El objetivo general de este campo es explorar y

construir sistemas capaces de emular o simular aspectos de la cognición humana, utilizando técnicas que van desde la lógica formal y los sistemas basados en reglas hasta las complejas redes neuronales inspiradas en el cerebro.

La idea de una «IA» monolítica y omnisciente es un producto de la ciencia ficción que no refleja la realidad tecnológica. Lo que existe en la práctica son los modelos de IA: programas de software altamente especializados, diseñados y entrenados para realizar tareas muy concretas. Un modelo de IA es, en esencia, una representación matemática o estadística de un problema específico, desarrollada a partir de grandes cantidades de datos (Google Cloud, s.f.a.). Su función es reconocer patrones, hacer predicciones o tomar decisiones dentro de un dominio estrictamente definido. Por ejemplo, un modelo puede ser experto en identificar rostros en fotografías, otro en traducir del español al inglés y un tercero en detectar transacciones fraudulentas. Cada uno de estos modelos es una herramienta distinta, con capacidades y limitaciones propias.

Para ilustrar esta distinción, podemos utilizar una analogía. Pensemos en la medicina como un campo de estudio vasto y

complejo. Dentro de la medicina, existen especialidades como la inmunología o la virología, que se centran en áreas más específicas del conocimiento médico. Un producto final derivado de esta investigación, como una vacuna contra la gripe, es una solución concreta, diseñada y probada para un problema muy específico: prevenir la infección por el virus de la influenza.

En esta analogía:

- el campo de la medicina representa la inteligencia artificial;
- una especialidad como la inmunología equivale a una subdisciplina, como el aprendizaje automático (*machine learning*);
- una vacuna específica corresponde a un modelo de IA.

Así como una vacuna contra la gripe es inútil para tratar una fractura ósea, un modelo de IA diseñado para generar imágenes no puede diagnosticar una enfermedad a partir de una radiografía, a menos que

haya sido específicamente entrenado para ello. Esta comprensión es fundamental, ya que permite analizar las capacidades, los sesgos y los riesgos de la IA no como un ente abstracto, sino a través de la lente de sus aplicaciones concretas y de sus modelos específicos.

CONTINUAR

2. ¿Cuáles son los principales fundamentos?

La IA comprende un ecosistema de subramas y conceptos clave interrelacionados. Resulta útil visualizarla como un conjunto amplio, dentro del cual existen subconjuntos más específicos. A continuación, se presentan las principales áreas y conceptos fundamentales:

2.1. Máquinas que aprenden: el universo del *machine learning* (ML)

El *machine learning* (ML) es un subconjunto específico de la IA enfocado en algoritmos que aprenden de los datos. En lugar de estar programados con reglas fijas, los sistemas de ML analizan grandes conjuntos de datos para identificar patrones y mejorar su desempeño con la experiencia. En términos simples, el ML permite que las computadoras tomen decisiones o hagan predicciones sin estar explícitamente programadas para cada caso, ajustando sus

parámetros internos conforme procesan más información (DataCamp, 2024).

Por ejemplo, un algoritmo de ML puede entrenarse con miles de imágenes etiquetadas para luego reconocer objetos en fotos nuevas. Su característica central es la adaptación automática: el modelo refina sus resultados a medida que recibe más datos y retroalimentación, volviéndose más preciso en sus predicciones (DataCamp, 2024).

IA vs. automatización basada en reglas

Cabe distinguir la «inteligencia artificial», especialmente las técnicas de «aprendizaje automático», de los sistemas automatizados tradicionales basados en reglas. En un sistema basado en reglas, el comportamiento está completamente definido por reglas explícitas escritas por programadores (por ejemplo, una serie de condiciones *if-then*); el sistema no «aprende» ni sale de lo previsto: solo ejecuta lo que sus reglas permiten (TechTarget, 2024). Este enfoque fue común en la primera época de la IA, así como en chatbots o programas de los años 60 y 70, como ELIZA, y en los denominados sistemas expertos de los 80. Su ventaja es la predictibilidad, pero presentan limitaciones: no pueden adaptarse a situaciones nuevas ni generalizar más allá de sus reglas programadas (TechTarget, 2024).

Por el contrario, en un sistema con «aprendizaje automático», el propio algoritmo infiere sus «reglas» a partir de los datos. Un modelo de ML ajusta sus parámetros internos con cada ejemplo que procesa, volviéndose capaz de manejar casos que no fueron previstos uno por uno por un programador (TechTarget, 2024). En suma, la IA moderna aporta flexibilidad y adaptación: donde una automatización fija colapsaría ante una condición no contemplada, un modelo entrenado con suficientes datos podría improvisar una respuesta razonable. Esta diferencia permite entender por qué la IA ha ganado terreno: supera el enfoque rígido de la programación tradicional, posibilitando soluciones para problemas complejos y variados, como el reconocimiento del habla, la conducción autónoma o la traducción idiomática, que sería impracticable resolver enumerando manualmente todas las reglas posibles.

Ejemplo de sistema basado en reglas: —

un sistema de alerta médica simple podría tener una regla como: Si la frecuencia cardíaca del paciente es > 120 lpm Y la presión arterial es $< 90/60$ mmHg, ENTONCES activar una alarma. Estos sistemas son predecibles y fiables, pero también rígidos. No pueden manejar situaciones que no fueron anticipadas en sus reglas y son difíciles de mantener a medida que la complejidad aumenta. Los «sistemas

expertos» de los años 80, como MYCIN para el diagnóstico de infecciones, son un ejemplo clásico de esta aproximación.

Ejemplo de sistema de machine learning (ML): —

un filtro de spam moderno es un excelente ejemplo de «*machine learning*». En lugar de que un programador escriba miles de reglas para identificar spam (por ejemplo, Si el correo contiene «oferta increíble», ENTONCES es spam), el sistema se «entrena» con un enorme conjunto de datos de correos electrónicos previamente etiquetados por humanos como «spam» o «no spam». El algoritmo de ML analiza estos ejemplos y aprende por sí mismo los patrones sutiles y complejos que distinguen un tipo de correo del otro. Esto le permite no solo ser más preciso, sino también adaptarse a nuevas tácticas de los spammers sin necesidad de ser reprogramado manualmente.

2.2 Profundizando en los datos: el poder del deep learning (DL)

El *deep learning* es un subconjunto dentro del *machine learning* que utiliza arquitecturas de redes neuronales artificiales de múltiples capas (de ahí lo de «profundo») para aprender de cantidades masivas de datos. Este enfoque permite descubrir patrones complejos y representaciones de alto nivel; imita, de cierta forma, el proceso de aprendizaje del cerebro humano mediante estas redes profundas.

En contraste con técnicas más simples de ML, los modelos de «deep learning» requieren generalmente muchos más datos y mayor poder de cómputo, pero logran un rendimiento superior en tareas como el reconocimiento de imágenes, de voz o de lenguaje natural (Coursera, 2023). En resumen, el «deep learning» es un tipo avanzado de aprendizaje automático que emplea redes neuronales profundas para extraer conocimiento y ha sido responsable de los mayores avances en IA en la última década, por ejemplo, en sistemas de traducción automática o en diagnóstico por imágenes médicas.

2.3 El cerebro artificial: redes neuronales artificiales (RNA)

Las redes neuronales artificiales (RNA) constituyen la base de la IA moderna, especialmente del aprendizaje profundo (*deep learning*). Una RNA es un modelo computacional inspirado en la estructura y el funcionamiento del cerebro humano. Consiste en un conjunto de nodos interconectados, denominados «neuronas artificiales», que se organizan en capas. Cada neurona recibe señales, realiza un cálculo simple y transmite el resultado a otras neuronas. Aunque la idea fue conceptualizada por primera vez en 1943 por Warren McCulloch y Walter Pitts, quienes propusieron que el cerebro podría modelarse como un sistema de unidades lógicas, su

verdadero potencial no se desató hasta la llegada de la computación moderna (IBM, 2023).

El «aprendizaje» en una red neuronal ocurre durante un proceso de entrenamiento, donde la red ajusta la fuerza de las conexiones entre sus neuronas para minimizar el error en sus predicciones. Este ajuste se realiza típicamente mediante un algoritmo llamado *backpropagation*.

Existen distintos tipos de RNA con distintas funcionalidades:

REDES NEURONALES CONVOLUCIONALES (CNN)

REDES NEURONALES RECURRENTE (RNN)

ARQUITECTURA TRANSFORMER

Son una arquitectura especializada en el procesamiento de datos que tienen una estructura de cuadrícula, como las imágenes. Su diseño se inspira en el córtex visual de los animales. Podemos imaginar una CNN como un equipo de especialistas que analiza una fotografía por etapas.

- **Capas convolucionales (los escáneres):** las primeras capas actúan como escáneres que utilizan «filtros» (pequeñas matrices de números) para deslizarse sobre la imagen y detectar características simples como bordes, esquinas, texturas o colores. Cada filtro se especializa en encontrar un patrón específico.
- **Capas de activación y agrupación (los resumidores):** después de cada escaneo, una capa de activación (comúnmente *ReLU*) decide qué características son lo suficientemente importantes

como para pasar a la siguiente etapa. Luego, una capa de agrupación (*pooling*) reduce el tamaño de la información, resumiendo los hallazgos de pequeñas regiones para hacer el proceso más eficiente y ayudar a que el modelo reconozca un objeto sin importar su posición en la imagen.

- **Capas totalmente conectadas (el detective jefe):** al final de la red, las capas totalmente conectadas reciben todas las características de alto nivel detectadas (por ejemplo, «tiene bigotes», «tiene orejas puntiagudas») y toman la decisión final, como clasificar la imagen con la etiqueta «gato».

Gracias a esta estructura jerárquica, las CNN son la tecnología detrás de la mayoría de los avances en visión por computadora, incluyendo el reconocimiento facial, la conducción autónoma y el diagnóstico médico por imágenes (Google Cloud, s.f.b.).

**REDES NEURONALES
CONVOLUCIONALES (CNN)**

**REDES NEURONALES
RECURRENTES (RNN)**

**ARQUITECTURA
TRANSFORMER**

Las RNN están diseñadas para trabajar con datos secuenciales, donde el orden de los elementos es importante, como el texto, el audio o las series temporales financieras. Su característica distintiva es un «bucle de retroalimentación» que permite que la información persista. En cada paso de la secuencia, la red no solo procesa la nueva entrada, sino que también considera la salida del paso anterior, dotándola de una forma de «memoria» a corto plazo.

Sin embargo, las RNN tradicionales presentan una limitación significativa conocida como el «problema del desvanecimiento del gradiente», que dificulta aprender y recordar dependencias en

secuencias largas. Para solucionarlo, se desarrollaron las redes de memoria a corto y largo plazo (LSTM). Las LSTM son un tipo avanzado de RNN que incorpora una estructura interna más compleja llamada «celda de memoria», controlada por tres «compuertas»:

- **Compuerta de olvido (*forget gate*).** Decide qué información del estado anterior de la celda es irrelevante y puede ser descartada.
- **Compuerta de entrada (*input gate*):** determina qué nueva información de la entrada actual es importante y debe ser almacenada en la celda.
- **Compuerta de salida (*output gate*):** controla qué parte de la información almacenada en la celda se utiliza para generar la salida en el paso actual.

Este sofisticado mecanismo de compuertas permite a las LSTM mantener y manipular una memoria a largo plazo, haciéndolas extremadamente efectivas para tareas como la traducción automática, el reconocimiento de voz y el análisis de sentimientos (NVIDIA, s.f.).

**REDES NEURONALES
CONVOLUCIONALES (CNN)**

**REDES NEURONALES
RECURRENTE (RNN)**

**ARQUITECTURA
TRANSFORMER**

Presentada por investigadores de Google en 2017, la arquitectura Transformer ha revolucionado el campo del procesamiento del lenguaje natural (NLP) y es la base de los modelos de lenguaje gigantes (LLM) como la serie GPT de OpenAI. Su innovación fundamental es el mecanismo de «autoatención» (*self-attention*).

A diferencia de las RNN, que procesan el texto palabra por palabra en secuencia, la autoatención permite al modelo sopesar la importancia de todas las demás palabras en una oración al interpretar una palabra específica. Por ejemplo, en la frase «El robot dejó caer la bola porque estaba oxidado», el mecanismo de atención puede determinar que «oxidado» se refiere a «el robot» y no a «la bola». Esta capacidad de capturar relaciones contextuales a larga distancia y de procesar todos los elementos de la secuencia en paralelo (en lugar de secuencialmente) ha permitido entrenar modelos mucho más grandes y potentes que sus predecesores, dando lugar a la era actual de la IA generativa (AWS Amazon, s.f.).

2.4 Inteligencia artificial generativa (IA generativa, IAG)

Es una rama emergente dentro del *deep learning*, enfocada en generar contenido nuevo y original. A diferencia de los sistemas discriminativos (que solo clasifican o reconocen patrones), los modelos generativos pueden crear textos, imágenes, audio o video que no existían previamente, basándose en los patrones aprendidos durante su entrenamiento (Oracle, 2025).

En otras palabras, la IA generativa extrapola a partir de sus datos de entrenamiento para producir ideas o

artefactos novedosos. Tecnologías recientes como los modelos de lenguaje gigantes (por ejemplo, GPT-3 y GPT-4 de OpenAI) y generadores de imágenes (por ejemplo, DALL-E, Midjourney) pertenecen a esta categoría. ChatGPT es un caso emblemático de IAG: un chatbot que puede conversar y producir respuestas coherentes en lenguaje natural, habiendo sido entrenado con billones de palabras de textos humanos.

La IA generativa se caracteriza por usar las mismas redes neuronales profundas, pero llevándolas un paso más allá para la creatividad artificial, pudiendo, por ejemplo, escribir un párrafo, componer música o diseñar un gráfico original a partir de una indicación humana (Red Hat, 2024).

CONTINUAR

3. Breve historia de la IA: hitos y evolución

La trayectoria histórica de la inteligencia artificial ha estado marcada por períodos de gran optimismo y por hitos notables, seguidos de fases de estancamiento y renovación. A continuación, se presenta un resumen cronológico de los acontecimientos clave en el desarrollo de la IA.

3.1 Los orígenes (1950-1960): el nacimiento de una idea revolucionaria

El punto de partida conceptual de la IA moderna se atribuye comúnmente al matemático británico Alan Turing. En su influyente artículo de 1950, *Computing Machinery and Intelligence*, Turing abordó la audaz pregunta: «¿Pueden las máquinas pensar?». Reconociendo la ambigüedad filosófica de la palabra «pensar», propuso un experimento de pensamiento pragmático para reemplazar la pregunta: el «Juego de la Imitación».

El juego original involucraba a tres participantes: un interrogador humano, un hombre y una mujer, todos en habitaciones separadas. El interrogador debía determinar, a través de preguntas escritas, cuál de los otros dos era la mujer. Turing adaptó este juego sustituyendo a uno de los participantes por una computadora. La nueva pregunta se convirtió en «¿podría una máquina conversar de manera tan convincente que un interrogador humano no pudiera distinguirla de un ser humano?». Si la respuesta era afirmativa, la máquina habría superado lo que hoy conocemos como el *test* de Turing, proporcionando una medida operativa de la inteligencia artificial.

Si Turing sentó las bases filosóficas, el nacimiento oficial de la IA como campo de investigación tuvo lugar en el verano de 1956 en el Dartmouth College (EE.UU.). Organizada por John McCarthy, Marvin Minsky, Nathaniel Rochester y Claude Shannon, la «Propuesta del Proyecto de Investigación de Verano de Dartmouth sobre Inteligencia Artificial» reunió a los pioneros del campo. Fue en esta propuesta donde McCarthy acuñó por primera vez el término «inteligencia artificial».

El evento se basó en una conjetura fundamental y optimista: «cada aspecto del aprendizaje o cualquier otra característica de la inteligencia puede, en principio, ser descrito con tanta

precisión que puede fabricarse una máquina para simularlo». Esta conferencia marcó el inicio de una era de entusiasmo y financiación, dando lugar a los primeros programas de IA, como el *Logic Theorist* de Newell y Simon, capaz de demostrar teoremas matemáticos.

Durante las décadas de 1950 y 1960 se desarrollaron los primeros programas de IA. Por ejemplo, en 1956 Allen Newell y Herbert Simon presentaron *Logic Theorist*, considerado el primer programa de IA capaz de demostrar teoremas lógicos, y en 1957 crearon *General Problem Solver*, un intento de algoritmo de propósito general para resolver problemas (IA Canarias, 2025). Estos sistemas iniciales mostraron que las máquinas podían imitar ciertos procesos del pensamiento humano, como seguir pasos lógicos. Sin embargo, también evidenciaron las grandes limitaciones tecnológicas de la época: el hardware disponible ofrecía poca potencia de cálculo y la falta de datos extensos impedía lograr algo más allá de problemas de juguete. A pesar del optimismo inicial, hacia finales de los años 60 se hizo patente que muchas promesas de la IA, como verdaderas conversaciones con máquinas o robots inteligentes, todavía estaban lejos de alcanzarse.

3.2 Los inviernos de la IA (1970-1980)

La brecha entre las expectativas y los resultados concretos llevó a un periodo de retracción conocido como el invierno de la IA. En los años 70, la investigación en IA sufrió un enfriamiento significativo. Las razones incluyeron que las computadoras no tenían suficiente poder de procesamiento ni memoria para cumplir las ambiciosas metas planteadas, y muchos proyectos no lograron avances tangibles (IA Canarias, 2025).

Gobiernos y agencias que antes financiaban generosamente la IA, como el Departamento de Defensa de EE.UU., recortaron sus apoyos tras informes técnicos pesimistas. Por ejemplo, el reporte Lighthill (Reino Unido, 1973) concluyó que la IA no había cumplido las expectativas. Consecuentemente, durante 1974-1980 la inversión y el interés en IA cayeron drásticamente (IA Canarias, 2025). Este primer invierno congeló muchos proyectos; pocos progresos destacables ocurrieron en ese periodo.

A mediados de los años 80 hubo un breve resurgir gracias a los sistemas expertos: programas basados en reglas que encapsulaban conocimiento de especialistas en dominios acotados, como MYCIN (1975) para diagnóstico médico. Los sistemas expertos demostraron utilidad en tareas específicas y revitalizaron la industria por algunos años (IA Canarias, 2025). No obstante, también tenían problemas: eran costosos

de desarrollar, difíciles de mantener y no se adaptaban fuera de su campo. Para fines de los 80, el mercado de sistemas expertos se saturó y nuevamente la confianza decayó. Se habla de un segundo «invierno de la IA» a finales de los 80 y principios de los 90, cuando el entusiasmo volvió a enfriarse debido a las limitaciones y al alto costo de estas soluciones rígidas (IA Canarias, 2025).

En síntesis, las décadas de 1970-80 fueron de avances lentos y frustración, enseñando a los investigadores que la inteligencia artificial era un desafío más complejo de lo que se pensó originalmente.

3.3 El renacimiento (1990-2000): el impulso del cómputo y los datos

A partir de los años 90, varios factores propiciaron un renacimiento de la IA. Uno de ellos fue la persistencia de ciertos grupos de investigación en redes neuronales (el enfoque conexionista que había quedado relegado en los 70) y en algoritmos de aprendizaje automático inspirados en estadísticas. Por otro lado, la creciente capacidad de las

computadoras personales y estaciones de trabajo en los 90 permitió ejecutar experimentos más ambiciosos.

Un momento decisivo llegó en 1997, cuando la supercomputadora de IBM *Deep Blue* venció al campeón mundial de ajedrez Garry Kasparov en un encuentro oficial (IBM, 2023). *Deep Blue* utilizaba técnicas de búsqueda avanzada y evaluación heurística, aunque no era aprendizaje automático puro, sino más bien un sistema especializado con enorme poder computacional. Aun así, este logro capturó la atención del público y demostró que la IA podía superar a un humano experto en una tarea compleja y bien definida. *Deep Blue* marcó un antes y después en la percepción de lo que la IA podía hacer, sacándola de los laboratorios y llevándola a titulares de prensa mundial.

A principios de los 2000, confluyeron avances que impulsaron fuertemente el aprendizaje automático (ML). Por un lado, mejoró el hardware: las computadoras eran mucho más rápidas y económicas que en las décadas previas, permitiendo entrenar modelos con más datos (DataCamp, 2024). Por otro lado, la explosión de datos digitales — originada por Internet, la web, sensores y bases de datos de empresas— ofreció la materia prima que los algoritmos de ML necesitaban para aprender. Algoritmos clásicos como las máquinas de soporte vectorial (SVM) o árboles de decisión,

junto con nuevas técnicas, empezaron a rendir muy bien cuando había suficientes datos disponibles. Durante los 2000, gigantes tecnológicos como Google, Amazon o Facebook invirtieron en IA para mejorar sus servicios: motores de búsqueda más inteligentes, recomendaciones personalizadas de productos o amigos, anuncios dirigidos y traducción automática básica, entre otros (DataCamp, 2024).

La IA comenzó a tener un impacto comercial real, integrándose en productos cotidianos, aunque muchas veces de forma tras bambalinas. Hitos adicionales incluyen la victoria del sistema IBM *Watson* en el concurso de preguntas *Jeopardy!* (2011), donde un modelo de preguntas y respuestas basado en aprendizaje automático derrotó a campeones humanos en un juego de conocimientos generales (IBM, 2023).

En resumen, durante las décadas de 1990 y 2000, la IA resurgió de sus cenizas, apoyada en algoritmos de aprendizaje más eficientes, mejor hardware y mucho más datos, preparando el terreno para la revolución que vendría después.

3.4 La revolución del *deep learning* (década de 2010)

El verdadero punto de inflexión en la historia moderna de la IA ocurrió en la década de 2010 con la llegada y consolidación del *deep learning* (aprendizaje profundo). Aunque el concepto de redes neuronales profundas existía desde antes, fue a partir de 2012 que demostraron un salto de rendimiento notable. En ese año, un modelo de *deep learning* (una red neuronal convolucional llamada AlexNet) ganó la competencia internacional ImageNet de reconocimiento de imágenes con una ventaja abrumadora sobre métodos tradicionales, reduciendo la tasa de error de ~25 % a ~16 % (Coursera, 2023).

Su éxito se debió principalmente al uso de GPU en su entrenamiento y a nuevas técnicas de *deep learning*. Este avance fue rápidamente seguido por mejoras en otros ámbitos: reconocimiento de voz con precisión casi humana, mejoras en sistemas de detección de rostros y capacidades de traducción automática neuronal mucho más fluidas. La combinación de grandes datos, gran poder de cómputo y nuevas técnicas de redes neuronales permitió que los algoritmos superaran barreras antes infranqueables (DataCamp, 2024).

Por ejemplo, en 2016, el programa AlphaGo de DeepMind derrotó al campeón mundial de Go, Lee Sedol, un juego exponencialmente más complejo que el ajedrez y

considerado por mucho tiempo un último bastión de la intuición humana. AlphaGo logró un 4-1 contra Sedol utilizando *deep learning* y aprendizaje por refuerzo para evaluar posiciones y planificar jugadas, asombrando a la comunidad global de expertos (IBM, 2023a). Este evento confirmó el poder del *deep learning* para abordar tareas de altísima complejidad estratégica.

Durante la década de 2010 se refinaron arquitecturas clave, como las redes neuronales convolucionales (CNN) para visión, las redes recurrentes (RNN y LSTM) para secuencias, y hacia finales de la década surgieron los ya mencionados *Transformers*, que resultaron extremadamente eficaces en tareas de lenguaje natural.

Gracias a estos avances, la IA comenzó a integrarse en productos de consumo masivo. Asistentes virtuales como Siri, Google Assistant o Alexa mejoraron sustancialmente en la comprensión del lenguaje hablado y de los contextos; los sistemas de recomendación de contenido (Netflix, Spotify, YouTube) se volvieron ubicuos; vehículos autónomos comenzaron pruebas en vías reales; y empresas de todos los sectores

incorporaron IA para análisis predictivo, mantenimiento preventivo, detección de fraudes, entre otros usos.

Hacia 2019, muchos ya hablaban de que estábamos en plena cuarta revolución industrial, impulsada en gran medida por la inteligencia artificial y el aprendizaje profundo (DataCamp, 2024).

3.5 La era generativa y la expansión masiva (años 2020)

En la actual década de 2020, la IA ha alcanzado una popularidad y expansión sin precedentes, principalmente gracias a los sistemas de IA generativa. El lanzamiento de modelos de lenguaje de gran tamaño (LLM, *large language models*) entrenados con cantidades colosales de datos textuales marcó un cambio cualitativo. OpenAI presentó *GPT-3* en 2020, demostrando una sorprendente capacidad para generar texto coherente y creativo.

Pero el gran catalizador fue la aparición de ChatGPT (OpenAI) a fines de 2022, una interfaz conversacional basada en GPT-3.5 que cualquier persona con conexión podía usar. En tan solo cinco días, ChatGPT alcanzó un millón de

usuarios, y en dos meses superó los 100 millones de usuarios, convirtiéndose en la aplicación de más rápido crecimiento de la historia (González, 2024). Esto marcó el inicio de una verdadera explosión mediática de la IA generativa: por primera vez, el público general experimentó de primera mano lo que una IA avanzada podía hacer, como escribir ensayos, responder preguntas complejas o asistir en programación.

Paralelamente, modelos generativos de imágenes, como DALL-E 2, Stable Diffusion y Midjourney —todos lanzados en 2022—, mostraron que las máquinas también podían crear arte visual a partir de descripciones textuales, alimentando tanto el asombro como el debate ético en torno a estas tecnologías.

La inteligencia artificial generativa se ha convertido así en el motor del auge actual de la IA. A diferencia de generaciones previas de IA, que se limitaban a analizar o clasificar datos, esta nueva ola crea contenido original extrapolando patrones aprendidos (Oracle, 2025).

Las aplicaciones son vastas: desde asistentes virtuales más conversacionales y humanos, hasta generación de código asistida por IA, como GitHub Copilot, pasando por contenido multimedia sintético, incluyendo imágenes, voces o videos

generados artificialmente. No obstante, esta fase también plantea retos importantes: las alucinaciones o errores de los modelos generativos, los dilemas sobre derechos de autor, sesgos en los datos que pueden reflejarse en las respuestas de la IA, y los posibles impactos en empleos, temas que se abordarán más adelante en este curso.

A pesar de ello, el consenso es que la IA está ya firmemente instalada como una tecnología transformadora en la sociedad contemporánea, y su historia sigue escribiéndose día a día con nuevos avances.

CONTINUAR

4. Análisis de la actualidad

Ahora que conocemos los fundamentos y la historia de la inteligencia artificial, resulta relevante realizar un análisis de la situación actual en la que nos encontramos y comprender qué nos permitió llegar hasta aquí.

4.1 ¿Qué factores impulsaron el auge actual de la IA?

Después de décadas de altibajos, ¿por qué la IA ha «explotado» en relevancia en los últimos años? Los fundamentos de este auge pueden resumirse en tres factores principales —dos tecnológicos y uno social—, además de ciertos desarrollos recientes en técnicas de IA.

**1. INCREMENTO DEL PODER
DE CÓMPUTO**

**2. GRANDES CANTIDADES DE
DATOS**

**3. ACCESIBILIDAD Y
DEMOCRATIZACIÓN DE LA IA**

El *hardware* disponible para desarrollar IA mejoró de manera exponencial. La Ley de Moore y los avances en ingeniería de semiconductores produjeron microprocesadores cada vez más rápidos y asequibles. Más importante aún, el uso de unidades de procesamiento gráfico (GPU) revolucionó el entrenamiento de modelos de *deep learning*.

Las GPU pueden realizar cálculos en paralelo de manera masiva, reduciendo dramáticamente el tiempo necesario para entrenar redes neuronales profundas, de semanas a días o incluso horas. Esta capacidad de procesamiento masivo era una pieza faltante en los primeros días de la IA, y su disponibilidad en la década de 2010 catalizó el desempeño de nuevos modelos (AWS, s.f.).

Sin el poder computacional moderno, incluida la computación en la nube a gran escala, los avances algorítmicos no habrían podido despegar. Hoy, en cambio, es factible entrenar modelos con miles de millones de parámetros gracias a la infraestructura computacional actual.

1. INCREMENTO DEL PODER DE CÓMPUTO

2. GRANDES CANTIDADES DE DATOS

3. ACCESIBILIDAD Y DEMOCRATIZACIÓN DE LA IA

La materia prima de la IA es la información. En los últimos 20-30 años, la humanidad ha generado volúmenes de datos sin precedentes; se estima que más del 90 % de todos los datos del mundo se produjeron tan solo en la última década. Este *big data* proviene de diversas fuentes: la web y las redes sociales, sensores de IoT, historiales de compras, registros médicos digitalizados, videos, entre otros.

Los algoritmos de *machine learning* y, especialmente, de *deep learning* requieren grandes cantidades de datos para alcanzar un alto rendimiento, y por fin, en la era contemporánea, esa condición está satisfecha (DataCamp, 2024). Por ejemplo, los modelos de visión artificial actuales se entrenan con millones de imágenes, y los modelos de lenguaje con trillones de palabras de texto. Esta abundancia de datos de entrenamiento ha permitido a las IA aprender representaciones más ricas del mundo real, logrando predicciones y generando resultados con mayor exactitud.

**1. INCREMENTO DEL PODER
DE CÓMPUTO**

**2. GRANDES CANTIDADES DE
DATOS**

**3. ACCESIBILIDAD Y
DEMOCRATIZACIÓN DE LA
IA**

El tercer factor es que la IA se volvió accesible para una amplia comunidad de desarrolladores, empresas e incluso usuarios comunes de Internet. Varias condiciones confluyeron en este cambio.

Primero, la aparición de *frameworks* de código abierto, como TensorFlow, PyTorch y Scikit-learn, y de bibliotecas públicas hizo que las herramientas de IA estuvieran disponibles para cualquier programador, sin necesidad de reinventar algoritmos desde cero. Segundo, muchos modelos preentrenados y publicaciones de investigación se comparten abiertamente, acelerando la innovación global. Tercero, los grandes proveedores de computación en la nube (AWS, Google Cloud, Azure) ofrecen servicios de IA preconstruidos o hardware especializado bajo demanda, eliminando la barrera de tener que poseer supercomputadoras propias.

Todo esto significa que hoy una startup pequeña o un grupo de estudiantes puede experimentar con IA avanzada con costos relativamente bajos. El enfoque de código abierto y el acceso en la nube han democratizado la IA, permitiendo que incluso pymes sin grandes presupuestos aprovechen modelos e infraestructura antes reservados a gigantes tecnológicos (Red Hat, 2025).

Además, la llegada de aplicaciones directas al usuario final, como apps con IA y chatbots públicos, hizo que millones de personas se familiaricen con la IA, impulsando su difusión viral. En conjunto, la facilidad para probar y adoptar soluciones de IA ha contribuido enormemente a su expansión actual.

Avances metodológicos recientes

Junto a los factores anteriores, vale la pena mencionar ciertos desarrollos técnicos dentro del campo de la IA que han potenciado su efectividad y seguridad, facilitando su adopción masiva en 2023-2025:

Aprendizaje por refuerzo con retroalimentación humana (RLHF): —

es una técnica de entrenamiento de modelos de IA que ha sido clave para mejorar la calidad de sistemas como ChatGPT. En RLHF, los modelos reciben *feedback* de evaluadores humanos sobre sus respuestas, y a partir de ello se entrena un *modelo de recompensa* que guía al agente de IA a optimizar su comportamiento mediante algoritmos de refuerzo (IBM, 2023a). En términos simples, los

humanos puntúan o comparan respuestas del modelo, y esas preferencias se usan para ajustar el modelo hacia resultados más deseables. Esto ha permitido que chatbots y asistentes de IA alineen mejor sus respuestas con las expectativas humanas, como ser útiles, veraces y evitar contenido inapropiado. El RLHF ha sido instrumental para hacer que las interacciones con IA generativa sean más seguras y útiles, reduciendo errores contextuales o salidas sin sentido mediante la incorporación de juicio humano en el ciclo de entrenamiento.

Generación aumentada por recuperación (RAG): —

es otra innovación reciente que mejora la precisión y actualidad de los modelos generativos. RAG combina un modelo de IA generativa con mecanismos de búsqueda en fuentes externas de información (Bismart, 2023). En lugar de depender únicamente del conocimiento fijo dentro de un modelo entrenado, un sistema RAG puede consultar una base de datos, documentos o Internet en tiempo real para obtener datos relevantes al usuario, e integrar esa información en su respuesta generativa. Esto soluciona en parte problemas como las «alucinaciones» o la desactualización de los LLM, ya que la IA puede fundamentar sus respuestas con datos externos verificables. Por ejemplo, un asistente con RAG podría buscar en una colección de artículos científicos el dato requerido y suministrar una respuesta apoyada en esa recuperación, combinando así lo mejor de dos mundos: el lenguaje natural generado por IA con la exactitud de una base de conocimiento específica.

Ecosistema open-source y colaboración: —

como se mencionó, el movimiento de código abierto en IA ha tenido un impacto fundamental. Modelos abiertos, como algunos de visión artificial o recientemente modelos de lenguaje de código abierto tipo LLaMA 2 de Meta, y comunidades colaborativas permiten que un amplio número de desarrolladores aporten mejoras, detecten errores y compartan resultados. Red Hat resume que el código abierto brinda soberanía tecnológica a empresas y países, y nivela el terreno para que *startups* y pymes compitan con grandes corporaciones en el uso de IA (Red Hat, 2025). Este ambiente colaborativo acelera la innovación y reduce el costo de entrada, contribuyendo también al auge de la IA a escala global.

4.2 La controversia en 2025 y perspectivas a futuro

03:26

Durante años, el progreso en la IA ha estado guiado por las «leyes de escalado» (*scaling laws*), una observación empírica que demostró que el rendimiento de los modelos mejoraba de forma predecible y consistente simplemente aumentando tres variables: el tamaño del modelo (más parámetros), la cantidad de datos de entrenamiento y la potencia de cómputo utilizada. Esta filosofía de «más grande es mejor» impulsó la carrera por construir LLM cada vez más gigantescos (Chen et al., 2025).

Sin embargo, en la actualidad existe un intenso debate en la comunidad científica sobre si este paradigma está llegando a sus límites. Investigaciones recientes y la experiencia con los modelos más avanzados sugieren que estamos entrando en una fase de rendimientos decrecientes. Duplicar el tamaño de un modelo ya no produce el mismo salto cualitativo en sus capacidades, un fenómeno conocido como «subescalado».

LAS POSIBLES CAUSAS DE ESTE ESTANCAMIENTO SON VARIAS:

- **Límites de los datos.** Se especula que podríamos estar agotando el suministro de datos textuales de alta calidad disponibles en Internet. A medida que los modelos se vuelven más grandes, la calidad, limpieza y diversidad de los datos se vuelven factores mucho más críticos que la simple cantidad bruta.
- **Límites arquitectónicos y de fiabilidad:** algunos estudios teóricos sugieren que las arquitecturas actuales, como el *Transformer*, pueden tener una tensión inherente entre su capacidad de aprendizaje y su fiabilidad. A mayor escala, estos modelos podrían volverse más propensos a la acumulación de

errores y a la generación de información espuria, limitando su utilidad para aplicaciones científicas o de alta precisión.

Esto no significa que el progreso de la IA se haya detenido, sino que podría estar cambiando de dirección. El futuro del campo probablemente no dependerá únicamente de construir modelos monolíticos cada vez más grandes, sino de la innovación en arquitecturas más eficientes, mejores datos y técnicas de entrenamiento, y nuevos paradigmas que permitan alcanzar una mayor inteligencia con un uso más sostenible de los recursos computacionales y energéticos. Entre estos nuevos cursos que pueden tomar los avances en IA en el futuro visible, podemos mencionar algunos en función de las últimas tendencias y capacidades de desarrollo actuales.

Agentes de IA autónomos —

A diferencia de los modelos actuales, que generalmente responden a una instrucción directa, se espera que los agentes de IA sean sistemas diseñados para comprender un objetivo, planificar los pasos necesarios y ejecutar tareas con mínima intervención humana.

Ya se están desarrollando sistemas con estas capacidades: pueden utilizar herramientas externas, acceder a una memoria para recordar interacciones pasadas y colaborar con otros agentes para resolver problemas complejos. En el futuro, estos agentes podrían actuar como «compañeros de trabajo digitales», automatizando operaciones empresariales complejas, desde la gestión de agendas hasta el análisis de datos.

De hecho, algunas proyecciones, como las de Gartner, sugieren que para 2028 una parte de las decisiones laborales podría ser tomada de forma autónoma por la IA.

IA multimodal —

La comunicación humana no se basa solo en texto. El futuro de la IA reside en su capacidad para comprender e integrar simultáneamente múltiples tipos de datos, como texto, imágenes, audio y video. Esta capacidad se conoce como «IA multimodal» y se perfila como la interfaz estándar para la interacción entre humanos y máquinas.

Ya se han observado avances notables en este campo, con modelos capaces de procesar y generar contenido a través de diversas modalidades. Al integrar distintos canales, la IA puede lograr una comprensión más completa del contexto, lo que permitiría interacciones más naturales e intuitivas.

Las aplicaciones prácticas ya se están explorando en sectores como la salud, donde los sistemas multimodales podrían combinar historiales médicos, imágenes y notas de voz para mejorar la precisión de los diagnósticos.

IA en el mundo físico —

Esta tendencia se manifiesta en dos áreas complementarias que ya están en desarrollo: la IA física y la *Edge AI* (IA en el borde).

La primera se refiere a la convergencia de la IA con la robótica y los dispositivos del *Internet of Things (IoT)*, lo que permite que los sistemas no solo procesen información, sino que actúen en el mundo real, como ya lo hacen los robots en logística o los drones autónomos.

Por otro lado, la IA en el borde implica trasladar el procesamiento desde la nube hacia los dispositivos locales, como teléfonos o sensores. Esto ofrece ventajas como respuestas en tiempo real y mayor privacidad de los datos, y ya se está implementando en áreas como la agricultura inteligente, donde permite optimizar el riego sin necesidad de una conexión constante a internet.

Nota aclaratoria sobre uso de IA

Este material fue asistido con herramientas de IA generativa para tareas de borrador, síntesis, reescritura y apoyo en la organización de contenidos. Cada sección fue revisada, editada y validada por el equipo humano, que verificó la precisión conceptual, la coherencia pedagógica y las fuentes citadas. Se invita a contrastar con las referencias bibliográficas incluidas y la documentación oficial. Dado que los modelos de IA evolucionan con rapidez, ciertas

especificaciones técnicas podrían actualizarse; este texto refleja el estado del conocimiento al momento de su elaboración.

CONTINUAR

Referencias

AWS. (s.f.). *¿Qué es una red neuronal?*
<https://aws.amazon.com/es/what-is-neural-network/>

Bismart. (2023). *¿Qué es un buscador inteligente RAG?*
<https://blog.bismart.com/que-es-buscador-inteligente-rag>

Coursera. (2023). *Deep learning vs. Machine learning: Guía para principiantes.* <https://www.coursera.org/mx/articles/ai-vs-deep-learning-vs-machine-learning-beginners-guide>

Chen, Z., Wang, S., Xiao, T., Wang, Y., Chen, S., Cai, X., He, J., & Wang, J. (2025). *Revisiting scaling laws for language models: The role of data quality and training strategies.* En *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Vol.1: Long Papers)* (pp.23881–23899). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/2025.acl-long.1163>

DataCamp. (2024). *¿Qué es el machine learning?*
<https://www.datacamp.com/es/blog/what-is-machine-learning>

González, V. (2024, 8 de enero). El 80% de las personas usan inteligencia artificial en su vida diaria y no lo saben. *Infobae*.
<https://www.infobae.com/educacion/2024/01/08/el-80-de-las-personas-usan-inteligencia-artificial-en-su-vida-diaria-y-no-lo-saben-afirma-la-unam/>

Google Cloud, (s.f.a.). *¿Qué es un modelo de IA?*
<https://cloud.google.com/discover/what-is-an-ai-model?hl=es-419>

Google Cloud. (s.f.b.). *¿Qué es una Red Neuronal Convolutiva?* <https://cloud.google.com/discover/what-are-convolutional-neural-networks?hl=es>

IA Canarias. (2025). *Historia y evolución de la inteligencia artificial.* <https://iacanarias.com/historia-y-evolucion-de-la-inteligencia-artificial/>

IBM. (2023). *Historia de la inteligencia artificial.*
<https://www.ibm.com/es-es/think/topics/history-of-artificial-intelligence>

IBM (2023a). *¿Qué es RLHF?* <https://www.ibm.com/es-es/think/topics/rhf>

NVIDIA. (s.f.). *Long short-term memory (LSTM).* <https://developer.nvidia.com/discover/lstm>

Oracle. (2025). *¿Qué es la IA generativa?* <https://www.oracle.com/mx/artificial-intelligence/generative-ai/what-is-generative-ai/>

Red Hat. (2024). *¿Qué es la inteligencia artificial generativa?* <https://www.redhat.com/es/topics/ai/what-is-generative-ai>

Tableau. (2023). *Qué es la inteligencia artificial: definición, historia, aplicaciones y futuro.* <https://www.tableau.com/es-mx/data-insights/ai/what-is>

TechTarget. (2024). *Choosing between a rule-based vs. machine learning system.* <https://www.techtarget.com/searchenterpriseai/feature/How-to-choose-between-a-rules-based-vs-machine-learning-system>

CONTINUAR

Descarga en PDF



Módulo 1. Fundamentos e historia de la inteligencia artificial.pdf

1.3 MB

