Investigación sobre Modelos de IA

PSWE-04 Diseño de Software

Daniel Canessa Valverde Stephanie Delgado Brenes

2025-07-18

Tabla de contenidos

Abstract

Este documento presenta un análisis estructurado de los modelos de inteligencia artificial (IA), abarcando desde sus fundamentos teóricos y principales tipos hasta el papel central de los modelos fundacionales en el estado del arte actual. Se examinan las aplicaciones más relevantes de la IA en ámbitos como la visión por computador, el procesamiento de lenguaje natural, la salud, las finanzas y la educación, así como los desafíos, riesgos y limitaciones asociados a su desarrollo y adopción. Finalmente, se discuten las tendencias emergentes y las perspectivas futuras del campo, poniendo especial énfasis en su impacto social y laboral, y en la necesidad de promover una inteligencia artificial ética, explicable y orientada al beneficio colectivo.

1 Introducción

La inteligencia artificial se ha convertido en una de las tecnologías más influyentes de nuestro tiempo, con un alcance que va mucho más allá del ámbito científico y tecnológico. El avance acelerado de sus modelos y aplicaciones está transformando la forma en se trabaja, apende, se consume información y se relacionan las personas, generando un impacto profundo en la sociedad y el entorno laboral.

Hoy en día, la IA no solo optimiza procesos industriales o facilita tareas administrativas, sino que está presente en sistemas de recomendación, servicios de salud, herramientas educativas, plataformas de comunicación y asistentes virtuales. Este despliegue masivo ha permitido la creación de nuevas oportunidades de empleo, el surgimiento de profesiones vinculadas al análisis y gestión de datos, así como una mayor personalización en servicios y productos. Al mismo tiempo, plantean importantes retos en cuanto a la automatización de tareas, la adaptación de las competencias laborales, la equidad en el acceso a la tecnología y la protección de derechos fundamentales como la privacidad y la no discriminación.

En los últimos años, el enfoque "AI first" se ha convertido en el nuevo mantra de muchas empresas y organizaciones líderes a nivel mundial. Esta estrategia implica situar la inteligencia artificial en el centro de la innovación y la toma de decisiones, priorizando el desarrollo y la integración de soluciones basadas en IA en todos los niveles de la organización. Adoptar una mentalidad "AI first" no solo redefine productos y servicios, sino que transforma modelos de negocio, procesos internos y la cultura misma del trabajo, acelerando la transición hacia entornos más automatizados, personalizados y centrados en los datos.

En este escenario de constante evolución, comprender el funcionamiento, los alcances y las limitaciones de los modelos de inteligencia artificial resulta esencial para anticipar sus efectos en la vida cotidiana, el mercado laboral y la organización social. Este documento ofrece un análisis estructurado sobre los principales enfoques y aplicaciones de los modelos de IA, junto con los desafíos y tendencias que determinarán su influencia en los próximos años.

2 Objetivos

2.1 Objetivo general

Analizar de manera estructurada los principales modelos de inteligencia artificial, sus aplicaciones actuales y su impacto en el ámbito social y laboral, identificando los desafíos, riesgos y tendencias que configuran el futuro de esta tecnología.

2.2 Objetivos específicos

- Describir los fundamentos teóricos y las principales tipologías de modelos de inteligencia artificial, resaltando sus diferencias, ventajas y limitaciones.
- Identificar y explicar los modelos fundacionales más relevantes en la actualidad, detallando sus características, áreas de aplicación y ejemplos concretos.
- Examinar los principales desafíos, limitaciones y riesgos asociados al uso y desarrollo de modelos de inteligencia artificial, especialmente en lo que respecta a la ética, la privacidad y la equidad.
- Analizar las tendencias emergentes y perspectivas futuras de la inteligencia artificial, considerando tanto las oportunidades de innovación como sus implicaciones sociales y laborales.

3 ¿Qué es la inteligencia artificial?

La inteligencia artificial (IA) se puede definir como la habilidad de que el software realice actividades que normalmente requieren inteligencia humana. En una definición más detallada, la IA es la capacidad de las máquinas para utilizar algoritmos, aprender de los datos y aplicar lo aprendido en la toma de decisiones, de forma similar a los seres humanos. La IA permite que los sistemas informáticos procesen grandes volúmenes de información, aprendan y ejecuten tareas complejas (Rouhiainen 2018).

Por otro lado, Russell y Norvig (2010) definen la inteligencia artificial como el estudio de los agentes inteligentes, es decir, sistemas que perciben su entorno y realizan acciones que maximizan sus posibilidades de éxito (Russell y Norvig 2010).

En ambos casos se enfatiza la importancia de la percepción y la acción orientada a objetivos, e incluyen áreas como el aprendizaje automático, la representación del conocimiento, el razonamiento y la resolución de problemas. La IA no solo busca replicar capacidades humanas, sino también desarrollar nuevas formas de inteligencia que pueden superar las limitaciones humanas en procesamiento, velocidad y precisión.

El desarrollo y alcance de la inteligencia artificial, así como sus principales modelos, aplicaciones y desafíos, se abordan en detalle en las siguientes secciones de este documento.

4 Modelos de inteligencia artificial

Un modelo de inteligencia artificial es un programa que ha sido entrenado con datos para identificar patrones, tomar decisiones o hacer predicciones sin intervención humana directa. Los modelos aplican algoritmos matemáticos a entradas de datos para generar salidas que resuelven tareas específicas, como la clasificación de imágenes, la traducción automática de documentos o la predicción de tendencias (Rouhiainen 2018; IBM 2024).

La diferencia entre un algoritmo y un modelo es fundamental: los algoritmos son procedimientos, generalmente descritos en lenguaje matemático o pseudocódigo, que se aplican a un conjunto de datos para cumplir una función o propósito determinado; mientras que el modelo es el resultado final de aplicar ese algoritmo al conjunto de datos, es decir, la representación entrenada que toma decisiones o realiza predicciones (IBM 2024).

Existen diferentes formas de clasificar los modelos de IA, según su objetivo, metodología o tipo de salida:

- **Por su enfoque metodológico:**
  - **Modelos generativos:** pueden crear nuevos datos, como imágenes, textos o música (IBM 2024).
  - **Modelos discriminativos:** se enfocan en distinguir o clasificar datos entre diferentes categorías, como ocurre en el diagnóstico médico automatizado (IBM 2024).
- **Por la naturaleza de la tarea:**
  - **Modelos de clasificación:** asignan etiquetas o categorías a los datos, por ejemplo, identificar correos como spam o no spam (IBM 2024).
  - **Modelos de regresión:** predicen valores numéricos continuos, por ejemplo, estimar precios de viviendas (IBM 2024).

La principal diferencia entre estos enfoques radica en cómo los modelos representan y procesan la relación entre los datos de entrada y salida, así como en el tipo de problema que buscan resolver.

En años recientes, el desarrollo de modelos fundacionales o preentrenados ha revolucionado el campo. Estos modelos, entrenados con grandes volúmenes de datos, pueden adaptarse a múltiples tareas específicas mediante técnicas de ajuste fino (fine-tuning) (IBM 2024; Bommasani et al. 2021).

Es importante señalar que los criterios de clasificación (objetivo, metodología, tipo de salida, etc) no son excluyentes ni jerárquicos, sino que ofrecen distintas perspectivas sobre cómo abordar el diseño y aplicación de los modelos. En la siguiente sección se cubriran los distintos tipos de modelos.

4.1 Tipos de modelos de inteligencia artificial

Se ha desarrollado una amplia gama de modelos para abordar distintos tipos de problemas. Estos modelos pueden clasificarse de varias maneras, pero a nivel académico, destacan los siguientes enfoques:

4.1.1 1. Modelos simbólicos o basados en reglas

Descripción:
Estos modelos representan el conocimiento mediante reglas lógicas y cadenas de inferencia del tipo "si… entonces…", permitiendo la automatización de la toma de decisiones en dominios acotados (Russell y Norvig 2010).

Historia y contexto:
Los modelos simbólicos o basados en reglas, también conocidos como "IA simbólica" o sistemas expertos, dominaron la primera etapa de la inteligencia artificial, especialmente entre las décadas de 1960 y 1980. Durante este periodo, la principal aproximación fue intentar representar el conocimiento humano y el razonamiento lógico a través de reglas explícitas y símbolos, lo que llevó al desarrollo de numerosos sistemas expertos para dominios específicos (Russell y Norvig 2010).

Ventajas:

- Son fáciles de interpretar y auditar, ya que el razonamiento es explícito (Russell y Norvig 2010).
- Permiten trazabilidad total en la toma de decisiones (Russell y Norvig 2010).
- Son útiles en dominios cerrados o donde el conocimiento es completamente explícito (Russell y Norvig 2010).

Limitaciones:

- No aprenden de los datos, dependen del conocimiento explícito programado por expertos, lo que genera un cuello de botella (Russell y Norvig 2010).
- Tienen dificultades para escalar a dominios complejos, pues el número de reglas puede crecer exponencialmente (Russell y Norvig 2010).
- Son poco flexibles ante nuevos problemas no previstos y tienen dificultades para manejar información incompleta (Russell y Norvig 2010).

Desarrollo posterior:
Aunque la popularidad de los sistemas simbólicos ha disminuido con el auge del aprendizaje automático, siguen siendo relevantes en ciertas aplicaciones industriales, legales y de control, y han sido integrados en sistemas híbridos que combinan reglas explícitas con modelos probabilísticos o conexionistas (Russell y Norvig 2010).

Ejemplo:
El sistema experto MYCIN para diagnóstico médico fue un referente en los años 70. Otros ejemplos notables incluyen DENDRAL (análisis químico) y XCON (configuración de sistemas informáticos) (Russell y Norvig 2010).

4.1.2 2. Modelos conexionistas (Redes neuronales artificiales)

Descripción:
Estos modelos están inspirados en cómo funciona el cerebro humano. Usan pequeñas unidades llamadas "neuronas artificiales" que se conectan entre sí para procesar información y aprender a partir de ejemplos (Russell y Norvig 2010; Goodfellow, Bengio, y Courville 2016).

Historia y contexto:
Las primeras redes neuronales artificiales se crearon en los años 1950 y 1960. Un modelo famoso de esa época es el perceptrón. Sin embargo, las redes más avanzadas y útiles empezaron a aparecer en los años 1980, cuando se inventaron métodos que permiten entrenar redes con varias capas. En los últimos años, este campo ha crecido mucho gracias al "deep learning", que permite resolver problemas complejos como reconocer imágenes, entender texto o traducir idiomas (Russell y Norvig 2010; Goodfellow, Bengio, y Courville 2016).

Ventajas:

- Son muy buenas para reconocer patrones y encontrar relaciones en datos como imágenes, sonidos o textos (Russell y Norvig 2010; Goodfellow, Bengio, y Courville 2016).
- Pueden aprender a partir de muchos ejemplos y luego aplicar ese aprendizaje a situaciones nuevas, siempre que los datos de entrenamiento sean variados y representativos (Goodfellow, Bengio, y Courville 2016).
- Funcionan bien en problemas donde hay muchos datos disponibles.

Limitaciones:

- Necesitan muchos datos y potencia de cómputo para entrenarse bien (Goodfellow, Bengio, y Courville 2016).
- Son modelos difíciles de entender por dentro, porque no se puede saber exactamente cómo llegan a una decisión. Esto puede ser un problema en áreas donde es importante explicar el motivo de cada decisión, como en la medicina o las finanzas (Goodfellow, Bengio, y Courville 2016).

- Si el modelo es demasiado complejo para la cantidad de datos que se tiene, puede aprender detalles innecesarios y equivocarse al enfrentarse a nuevos casos (Russell y Norvig 2010).

Desarrollo posterior:
Las redes neuronales han dado lugar a muchos tipos de modelos. Por ejemplo, las redes convolucionales (CNN) se usan para analizar imágenes y las redes recurrentes (RNN) para trabajar con datos en secuencia, como texto o audio. Más recientemente, modelos como los transformers han hecho posible grandes avances en procesamiento de lenguaje natural y otras áreas (Goodfellow, Bengio, y Courville 2016).

Ejemplo:
Las redes neuronales convolucionales (CNN) permiten que los celulares reconozcan caras en fotos. Las redes recurrentes (RNN) se usan para el reconocimiento de voz y la traducción automática de textos. Los transformers están detrás de sistemas como ChatGPT (Goodfellow, Bengio, y Courville 2016).

4.1.3 3. Modelos probabilísticos

Descripción:
Estos modelos usan la estadística y las probabilidades para tomar decisiones cuando hay incertidumbre o información incompleta. Permiten que la inteligencia artificial maneje situaciones donde no se tiene toda la información o donde el estado puede cambiar (Russell y Norvig 2010; Bishop 2006).

Historia y contexto:
Los modelos probabilísticos comenzaron a usarse en inteligencia artificial cuando se vio que las reglas y la lógica no eran suficientes para tratar con el mundo real (este es incierto y muchas veces impredecible). Por eso, se empezaron a aplicar técnicas de probabilidad y estadística para ayudar a las máquinas a razonar bajo incertidumbre (Russell y Norvig 2010).

Ventajas:

- Pueden combinar conocimiento previo (por ejemplo, experiencias pasadas) con nueva información (Russell y Norvig 2010; Bishop 2006).
- Son muy útiles en tareas como diagnóstico médico, donde a veces no se tienen todos los datos (Russell y Norvig 2010).
- Permiten calcular qué tan probable es que ocurra algo, y tomar decisiones con base en esas probabilidades (Bishop 2006).

Limitaciones:

- Los modelos complejos pueden ser difíciles de construir y requieren muchos datos para ajustar bien sus parámetros (Bishop 2006).
- No siempre es fácil obtener las probabilidades exactas necesarias para el modelo (Russell y Norvig 2010).

- Su rendimiento puede bajar si los datos son muy ruidosos o poco representativos (Russell y Norvig 2010).

Desarrollo posterior:
Se han creado muchos tipos de modelos probabilísticos, como las redes bayesianas y los modelos ocultos de Markov. Estos se usan en áreas como reconocimiento de voz, diagnóstico de enfermedades y análisis de texto (Bishop 2006).

Ejemplo:
Una red bayesiana puede ayudar a un médico a estimar la probabilidad de una enfermedad, tomando en cuenta síntomas observados y experiencias previas. Los modelos ocultos de Markov se utilizan, por ejemplo, en el reconocimiento de voz para determinar qué palabras está diciendo una persona (Russell y Norvig 2010; Bishop 2006).

4.1.4 4. Modelos evolutivos

Descripción:
Estos modelos imitan el proceso de la evolución natural. Utilizan técnicas como la selección, la combinación y la mutación de soluciones para encontrar la mejor respuesta a un problema. En lugar de buscar una única solución desde el principio, generan muchas posibles soluciones y mejoran a través de ciclos de "prueba y error" (Russell y Norvig 2010).

Historia y contexto:
Los modelos evolutivos surgieron inspirados por las ideas de Charles Darwin sobre la evolución y la selección natural. Se han usado desde la década de 1960 en la inteligencia artificial, especialmente para resolver problemas de optimización en los que otras técnicas no funcionan bien (Russell y Norvig 2010).

Ventajas:

- Son muy buenos para buscar soluciones óptimas en problemas complejos o donde hay muchas variables posibles (Russell y Norvig 2010).
- No requieren que el problema tenga una fórmula matemática clara o sencilla (Russell y Norvig 2010).
- Pueden adaptarse a cambios en el entorno o en los requisitos del problema (Russell y Norvig 2010).

Limitaciones:

- El proceso puede ser lento y requerir mucho poder de cómputo, ya que se prueban muchas soluciones (similar al "backtracking") (Russell y Norvig 2010).
- No siempre garantizan encontrar la mejor solución posible, solo una solución "suficientemente buena" (Russell y Norvig 2010).
- El resultado puede depender mucho de cómo se define el proceso de selección y mutación (Russell y Norvig 2010).

Desarrollo posterior:
Con el tiempo, los modelos evolutivos se han combinado con otras técnicas de inteligencia artificial, como redes neuronales o modelos probabilísticos, para aprovechar las ventajas de ambos enfoques. Se siguen utilizando en la optimización de diseños, la ingeniería y la generación automática de soluciones (Russell y Norvig 2010).

Ejemplo:
Los algoritmos genéticos, que simulan el cruce y la mutación de genes, se emplean para encontrar la mejor forma de diseñar un circuito electrónico, planificar rutas de transporte, o resolver rompecabezas complejos (Russell y Norvig 2010).

4.1.5 5. Modelos de aprendizaje automático (Machine Learning)

Descripción:
Los modelos de aprendizaje automático pueden aprender a partir de los datos, sin necesidad de que una persona les programe todas las reglas. Su objetivo es mejorar en una tarea específica utilizando ejemplos previos o experiencias pasadas (Russell y Norvig 2010; Bishop 2006).

Historia y contexto:
El aprendizaje automático comenzó a desarrollarse como una forma de que las computadoras pudieran reconocer patrones y tomar decisiones por sí mismas, no solo siguiendo instrucciones fijas. Ha crecido mucho gracias al aumento de datos disponibles y la mejora en la potencia de cómputo. Hoy en día, machine learning es central en aplicaciones como motores de búsqueda, reconocimiento de voz y recomendaciones personalizadas (Russell y Norvig 2010).

Ventajas:

- Se adaptan y mejoran a medida que procesan más datos (Bishop 2006).
- Permiten resolver problemas complejos donde sería muy difícil o imposible escribir reglas manualmente.
- Son versátiles, ya que pueden aplicarse en tareas de clasificación, predicción, agrupamiento y detección de patrones (Bishop 2006).

Limitaciones:

- Requieren datos de buena calidad y en cantidad suficiente para aprender correctamente.
- Si los datos están sesgados, el modelo puede aprender esos sesgos y cometer errores (Bishop 2006).
- Los modelos complejos pueden ser difíciles de interpretar y explicar, especialmente para quienes no son expertos (Bishop 2006).

Desarrollo posterior:
Existen tres categorías dentro del aprendizaje automático (Russell y Norvig 2010; Bishop 2006):

- **Aprendizaje supervisado:** El modelo aprende con datos etiquetados, es decir, cada ejemplo tiene una respuesta correcta. Se usa mucho para clasificación (por ejemplo, saber si un correo es spam o no) y para predecir valores numéricos (regresión).
- **Aprendizaje no supervisado:** El modelo busca patrones o estructuras en datos sin etiquetas. Se emplea para agrupar datos similares o para reducir la cantidad de variables (reducción de dimensionalidad).
- **Aprendizaje por refuerzo:** El modelo aprende tomando decisiones, probando diferentes acciones y recibiendo recompensas o castigos. El objetivo es aprender la mejor estrategia posible a largo plazo.

Ejemplo:

- El filtrado de correos electrónicos spam (aprendizaje supervisado).
- La agrupación de clientes según su comportamiento de compra (aprendizaje no supervisado).
- Sistemas que aprenden a jugar videojuegos mediante prueba y error (aprendizaje por refuerzo).

4.1.6 6. Modelos generativos y discriminativos

Descripción:
Estos son dos enfoques diferentes para que la inteligencia artificial aprenda a partir de los datos:

- **Modelos generativos:** aprenden a comprender cómo se distribuyen y relacionan los datos de entrada y salida; incluso pueden generar nuevos datos parecidos a los que han visto (Goodfellow, Bengio, y Courville 2016; Bishop 2006).
- **Modelos discriminativos:** se enfocan únicamente en distinguir o clasificar los datos, es decir, en aprender la diferencia entre categorías o clases, sin preocuparse por cómo se generan los datos (Goodfellow, Bengio, y Courville 2016; Bishop 2006).

Historia y contexto:
El concepto de modelos generativos y discriminativos surgió para comparar métodos de aprendizaje automático que pueden, o no, generar nuevos ejemplos.

- Los modelos generativos, como Naive Bayes, existen desde hace mucho y se usaban sobre todo en problemas de clasificación de texto. Con el avance del "deep learning", aparecieron modelos generativos capaces de crear imágenes, textos y sonidos.
- Los modelos discriminativos se popularizaron porque suelen ser más eficaces cuando la tarea es solo clasificar (Goodfellow, Bengio, y Courville 2016).

Ventajas:

- **Generativos:** Permiten crear nuevos datos, pueden funcionar aunque haya pocos ejemplos etiquetados y son útiles para tareas como la generación de imágenes, texto o música (Goodfellow, Bengio, y Courville 2016).
- **Discriminativos:** Suelen tener mejor rendimiento en tareas de clasificación, ya que se concentran en distinguir entre categorías (Goodfellow, Bengio, y Courville 2016).

Limitaciones:

- **Generativos:** Son más complejos de entrenar y requieren mucho poder de cómputo, especialmente en modelos modernos (Goodfellow, Bengio, y Courville 2016).
- **Discriminativos:** No pueden generar datos nuevos y dependen de la disponibilidad de ejemplos bien etiquetados (Goodfellow, Bengio, y Courville 2016).

Desarrollo posterior:
La investigación reciente ha llevado al desarrollo de modelos generativos cada vez más potentes, como los modelos de difusión y las GANs ("Generative Adversarial Networks"), que pueden crear imágenes realistas o textos coherentes. También han surgido modelos que combinan lo mejor de ambos enfoques para aprovechar sus ventajas en diferentes tareas (Goodfellow, Bengio, y Courville 2016).

Ejemplo:

- Un **modelo generativo** como una GAN puede crear imágenes nuevas de personas que no existen (Goodfellow, Bengio, y Courville 2016; Bishop 2006).
- Un **modelo discriminativo** como la regresión logística o una red neuronal profunda puede clasificar correos electrónicos como spam o no spam (Goodfellow, Bengio, y Courville 2016; Bishop 2006).

4.1.7 7. Modelos fundacionales (Foundation Models)

Descripción:
Son modelos de aprendizaje profundo que han sido entrenados previamente con grandes cantidades de datos variados (textos, imágenes, audio) y que pueden adaptarse a muchas tareas diferentes mediante ajustes específicos (Bommasani et al. 2021).

Historia y contexto:
Los modelos fundacionales surgieron en los últimos años como una evolución de los modelos de "deep learning" tradicionales. Antes, se entrenaba un modelo desde cero para cada tarea, ahora, gracias al poder de cómputo y la disponibilidad de datos, se entrena un solo modelo muy grande (como GPT) y luego se adapta a diferentes aplicaciones específicas. Esto ha permitido avances muy rápidos en áreas como el procesamiento de lenguaje natural, la generación de texto y la visión por computador (Bommasani et al. 2021).

Ventajas:

- Son muy versátiles: el mismo modelo puede servir para tareas muy distintas entre sí como resumir textos, responder preguntas, traducir idiomas o generar imágenes (Bommasani et al. 2021).
- Permiten aprovechar el conocimiento aprendido de millones de ejemplos en múltiples dominios (Bommasani et al. 2021).
- Pueden alcanzar resultados con menos datos y ajustes para cada tarea específica (Bommasani et al. 2021).

Limitaciones:

- Su entrenamiento inicial requiere una enorme cantidad de recursos computacionales, energía y tiempo (Bommasani et al. 2021).
- Al ser tan grandes, su funcionamiento interno es difícil de entender (Bommasani et al. 2021).
- Pueden aprender o amplificar sesgos presentes en los datos de entrenamiento, lo que puede afectar la equidad de sus resultados (Bommasani et al. 2021).

Desarrollo posterior:
El auge de los modelos fundacionales ha impulsado el desarrollo de nuevas aplicaciones y la integración de IA en productos de uso cotidiano. La tendencia actual es crear modelos aún más grandes y potentes, así como investigar cómo hacerlos más justos, eficientes y explicables.

Ejemplo:

- GPT-4 de OpenAI (que puede generar textos y programar) (Bommasani et al. 2021).
- Llama de Meta (que puede adaptarse a tareas de lenguaje en muchos idiomas) (Bommasani et al. 2021).

Además, existen enfoques híbridos y emergentes, como los sistemas neuro-simbólicos, que buscan combinar la robustez del razonamiento lógico con la capacidad de aprendizaje de los modelos conexionistas (Russell y Norvig 2010).

4.1.8 Resumen de tipos de modelos

La siguiente tabla resume los distintos tipos de modelos de inteligencia artificial:

Tabla 4.1: Tabla 1. Resumen de los principales tipos de modelos de inteligencia artificial.

Tipo de modelo	Descripción breve	Ventajas principales	Limitaciones principales	Ejemplo destacado
Simbólicos / basados en reglas	Usan reglas lógicas explícitas para tomar decisiones	Fáciles de interpretar y auditar. Útiles en dominios cerrados	No aprenden de datos. Difíciles de escalar y poco flexibles	MYCIN (diagnóstico médico)
Conexionistas (redes neuronales)	Aprenden a partir de ejemplos usando "neuronas" artificiales	Excelentes en reconocimiento de patrones y manejo de grandes volúmenes de datos	Necesitan muchos datos y cómputo. Difíciles de interpretar ("cajas negras")	CNN para reconocimiento de imágenes
Probabilísticos	Usan la estadística y probabilidad para razonar bajo incertidumbre	Combinan conocimiento previo y datos nuevos. Útiles en tareas con incertidumbre	Difíciles de construir y ajustar. Requieren estimar probabilidades precisas	Redes bayesianas en diagnóstico médico
Evolutivos	Imitan la evolución natural para optimizar soluciones	Ideales para problemas complejos o con muchas variables	Proceso lento. No garantizan la mejor solución posible	Algoritmos genéticos en diseño de circuitos
Aprendizaje automático (ML)	Aprenden patrones automáticamente a partir de datos	Se adaptan y mejoran con experiencia. Versátiles en muchas tareas	Requieren datos de calidad. Pueden aprender sesgos o ser difíciles de explicar	Filtrado de spam en correo electrónico
Generativos	Aprenden la distribución de los datos y pueden generar ejemplos nuevos	Permiten crear datos, útiles con pocos ejemplos etiquetados	Complejos de entrenar. Requieren mucho cómputo	GANs generando imágenes
Discriminativos	Se enfocan en clasificar o distinguir entre categorías	Eficaces en clasificación. Suelen requerir menos cómputo	No pueden generar datos nuevos. Necesitan datos etiquetados	Regresión logística para spam/no spam
Fundacionales	Modelos preentrenados muy grandes y versátiles para muchas tareas	Muy versátiles. Aprenden de muchos dominios y tareas	Requieren enormes recursos. Difíciles de explicar. Pueden amplificar sesgos	GPT-4, BERT, Llama

Tipo de modelo	Descripción breve	Ventajas principales	Limitaciones principales	Ejemplo destacado
Híbridos / emergentes	Combinan diferentes enfoques (ej: reglas + redes neuronales)	Buscan combinar lo mejor de varios métodos	Pueden ser complejos de diseñar e interpretar	Sistemas neuro-simbólicos

4.2 Modelos fundacionales: ejemplos y principales aplicaciones

Como se describió en la sección anterior, los **modelos fundacionales** han transformado la inteligencia artificial por su capacidad de aprender de grandes volúmenes de datos y adaptarse a múltiples tareas mediante ajuste fino (Bommasani et al. 2021).

La mayoría de los modelos fundacionales se basan en la arquitectura "transformer", que aporta la capacidad de aprender sobre relaciones complejas utilizando grandes cantidades de datos (Bommasani et al. 2021; Goodfellow, Bengio, y Courville 2016). Existen varios tipos de arquitectura transformer:

- **Transformer autoregresivo:** En sus salidas genera una palabra (o elemento) a la vez, usando la información previa de la secuencia, por ejemplo: GPT-4.
- **Transformer bidireccional:** Analiza simultáneamente el contexto previo y posterior de cada palabra, lo que mejora la comprensión del significado en una frase, por ejemplo: BERT.
- **Transformer multimodal:** Puede procesar diferentes tipos de datos a la vez, como texto, imágenes, audio o video, integrando información de varias fuentes, por ejemplo: Gemini.

A continuación se listan algunos de los modelos fundacionales más importantes de la actualidad:

1. GPT-4 (OpenAI)

- **Categoría:** Lenguaje natural generativo
- **Descripción:** Modelo avanzado especializado en la generación y comprensión de texto, capaz de responder preguntas, traducir, resumir y programar en varios lenguajes.
- **Tipo de arquitectura:** Transformer autoregresivo
- **¿Código abierto?:** No
- **Versiones principales:** GPT-1, 2, 3, 3.5, 4
- **Año:** 2023
- **Referencia:** (OpenAI 2023)

2. GPT-4o (OpenAI)

- **Categoría:** Lenguaje natural generativo y multimodal
- **Descripción:** Modelo que permite entrada y salida en texto, imagen y audio, con mejoras significativas en velocidad, eficiencia y capacidad multimodal respecto a GPT-4.
- **Tipo de arquitectura:** Transformer autoregresivo multimodal
- **¿Código abierto?:** No
- **Año:** 2024
- **Referencia:** (OpenAI 2024)

3. Claude 3 Opus (Anthropic)

- **Categoría:** Lenguaje natural generativo y multimodal
- **Descripción:** Modelo avanzado de lenguaje que destaca en razonamiento, comprensión de textos extensos y capacidades multimodales.
- **Tipo de arquitectura:** Transformer multimodal
- **¿Código abierto?:** No
- **Versiones principales:** Claude 3 Opus, Sonnet, Haiku
- **Año:** 2024
- **Referencia:** (Anthropic 2024)

4. Gemini 1.5 Pro (Google DeepMind)

- **Categoría:** Modelo multimodal
- **Descripción:** Modelo que integra texto, imagen, audio y video, con capacidad para contexto extenso y razonamiento complejo en múltiples dominios.
- **Tipo de arquitectura:** Transformer multimodal
- **¿Código abierto?:** No
- **Versiones principales:** Gemini 1.0, Gemini 1.5
- **Año:** 2024
- **Referencia:** (DeepMind, OpenAI, et al. 2023)

5. LLaMA 3 (Meta)

- **Categoría:** Lenguaje natural generativo
- **Descripción:** Modelo abierto y eficiente, entrenado con una gran cantidad de textos en varios idiomas y diseñado para aplicaciones de generación y comprensión de texto.
- **Tipo de arquitectura:** Transformer autoregresivo
- **¿Código abierto?:** Sí
- **Versiones principales:** LLaMA 1, 2, 3
- **Año:** 2024
- **Referencia:** (Meta AI 2024)

6. Mistral Mixtral 8x7B (Mistral AI)

- **Categoría:** Lenguaje natural generativo

- **Descripción:** Modelo open-source basado en arquitectura Mixture-of-Experts, optimizado para eficiencia y alto desempeño en tareas de lenguaje y código.
- **Tipo de arquitectura:** Transformer autoregresivo
- **¿Código abierto?:** Sí
- **Año:** 2023
- **Referencia:** (Mistral AI 2023)

7. Command R (Cohere)

- **Categoría:** Recuperación aumentada por generación (RAG)
- **Descripción:** Modelo diseñado para integración de búsqueda eficiente y generación de texto en contextos empresariales y conversacionales.
- **Tipo de arquitectura:** Transformer autoregresivo
- **¿Código abierto?:** No
- **Año:** 2024
- **Referencia:** (Cohere 2024)

8. Grok (xAI)

- **Categoría:** Lenguaje natural generativo
- **Descripción:** Modelo enfocado en integración de información en tiempo real y razonamiento contextual avanzado, con acceso directo a información de redes sociales.
- **Tipo de arquitectura:** Transformer autoregresivo
- **¿Código abierto?:** Parcialmente
- **Año:** 2023
- **Referencia:** (xAI 2023)

9. CLIP (OpenAI)

- **Categoría:** Visión por computador y lenguaje
- **Descripción:** Modelo pionero en relacionar imágenes y texto, usado ampliamente en aplicaciones multimodales de visión-lenguaje.
- **Tipo de arquitectura:** Transformer multimodal
- **¿Código abierto?:** Sí
- **Año:** 2021
- **Referencia:** (Radford et al. 2021)

10. Segment Anything Model (SAM) (Meta)

- **Categoría:** Visión por computador (segmentación de imágenes)
- **Descripción:** Modelo universal para segmentar cualquier objeto en una imagen, permitiendo anotación automática de objetos sin necesidad de entrenamiento adicional.
- **Tipo de arquitectura:** Transformer para visión
- **¿Código abierto?:** Sí
- **Año:** 2023

- **Referencia:** (Kirillov et al. 2023)

4.3 Principales aplicaciones de los modelos fundacionales

La siguiente tabla presentan ejemplos de áreas de aplicación de los modelos fundacionales:

Tabla 4.2: Tabla 2. Modelos fundacionales aplicados en diferentes áreas.

Área de aplicación	Descripción	Modelos fundacionales destacados
Visión por computador	Análisis y comprensión de imágenes, segmentación, búsqueda visual	CLIP, Segment Anything Model (SAM), GPT-4o, Gemini 1.5 Pro
Procesamiento de lenguaje natural	Generación y comprensión de texto, traducción, chatbots, análisis semántico	GPT-4, GPT-4o, Claude 3 Opus, Gemini 1.5 Pro, LLaMA 3, Mistral Mixtral 8x7B, Command R, Grok
Planificación y control	Robótica autónoma, optimización de rutas, juegos y simulaciones	Gemini 1.5 Pro, GPT-4o, Claude 3 Opus
Recomendación y personalización	Sistemas de recomendación y personalización de contenido	GPT-4o, Claude 3 Opus, Gemini 1.5 Pro, Command R
Salud	Análisis de imágenes clínicas, apoyo en diagnóstico, predicción de riesgos	Segment Anything Model (SAM), CLIP, Gemini 1.5 Pro, GPT-4o
Finanzas	Análisis y predicción de mercados, detección de fraudes, “scoring” crediticio	GPT-4, Claude 3 Opus, LLaMA 3, Mistral Mixtral 8x7B, Command R
Educación	Tutores inteligentes, generación de materiales, personalización de aprendizaje	GPT-4o, Gemini 1.5 Pro, Claude 3 Opus, LLaMA 3, Mistral Mixtral 8x7B

5 Desafíos, limitaciones y riesgos

A pesar de los grandes avances de la inteligencia artificial y los modelos fundacionales, existen importantes desafíos y riesgos que deben considerarse en su desarrollo y adopción (Bommasani et al. 2021; Russell y Norvig 2010):

- **Sesgo y discriminación:** Los modelos pueden perpetuar o incluso amplificar sesgos presentes en los datos de entrenamiento, lo que puede llevar a decisiones injustas o resultados discriminatorios en ámbitos como la salud o la contratación (Bommasani et al. 2021).
- **Explicabilidad y transparencia:** Muchos modelos modernos, especialmente los basados en "deep learning", funcionan como "cajas negras", dificultando la comprensión de cómo se toman las decisiones y limitando su adopción en sectores regulados o críticos (Goodfellow, Bengio, y Courville 2016).
- **Privacidad y seguridad de los datos:** El uso de grandes volúmenes de datos personales o sensibles plantea riesgos de privacidad, así como posibles vulnerabilidades ante ataques de manipulación o extracción de información (Bommasani et al. 2021).
- **Requerimientos computacionales y ecológicos:** El entrenamiento de modelos fundacionales requiere enormes recursos computacionales, lo que implica altos costos económicos y un impacto ambiental considerable debido al consumo energético (Bommasani et al. 2021).
- **Riesgo de desinformación y mal uso:** Los modelos generativos pueden crear información falsa difícil de distinguir de la real, lo que representa un riesgo para la integridad informativa y la confianza social (Bommasani et al. 2021).
- **Dependencia tecnológica y concentración de poder:** El desarrollo y control de los modelos más avanzados está en manos de pocas empresas, lo que puede limitar la innovación abierta y aumentar la desigualdad en el acceso a la tecnología (Bommasani et al. 2021).

6 Perspectivas futuras y tendencias

El campo de la inteligencia artificial y los modelos fundacionales sigue evolucionando rápidamente. Entre las principales tendencias y perspectivas a futuro destacan (Bommasani et al. 2021; Goodfellow, Bengio, y Courville 2016; Russell y Norvig 2010):

- **Avances en modelos multimodales:** Se espera que los modelos sean cada vez más capaces de integrar información de texto, imagen, audio y video, permitiendo aplicaciones más completas y contextuales.
- **Mayor eficiencia y sostenibilidad:** Habrá un énfasis en el desarrollo de modelos más eficientes, que requieran menos recursos para entrenarse y operar, buscando minimizar el impacto ambiental y ampliar el acceso a la tecnología.
- **Modelos explicables y auditables:** Crece la demanda de sistemas que permitan entender y controlar mejor las decisiones de la IA, facilitando su uso en áreas sensibles y reguladas.
- **Personalización y adaptación dinámica:** Los modelos tenderán a adaptarse mejor a contextos locales y necesidades individuales, mejorando la experiencia del usuario y la relevancia de sus aplicaciones.
- **Colaboración humano-IA:** Se espera una integración más estrecha entre humanos y sistemas inteligentes, donde la IA actúe como herramienta de apoyo y amplificadora de capacidades, más que como reemplazo directo.
- **Descentralización y democratización de la IA:** Proyectos open-source y modelos más accesibles facilitarán la innovación y el uso responsable en diversos contextos globales.
- **Fortalecimiento de regulaciones y ética:** Los marcos regulatorios y las consideraciones éticas serán cada vez más relevantes para guiar el desarrollo y despliegue de sistemas de IA seguros, justos y confiables.

7 Conclusiones

- El análisis de los fundamentos teóricos y de los principales tipos de modelos de inteligencia artificial evidencia la riqueza y diversidad del campo. Desde los modelos simbólicos y conexionistas hasta los probabilísticos, evolutivos, de aprendizaje automático y fundacionales, cada enfoque aporta distintas capacidades y limitaciones, adaptándose a contextos y problemas específicos. La clasificación de los modelos según su metodología, tarea o arquitectura permite comprender mejor tanto su potencial como sus restricciones, facilitando una selección informada para cada aplicación.
- La aparición de modelos fundacionales ha revolucionado el panorama de la inteligencia artificial, permitiendo una flexibilidad y escalabilidad inéditas. Modelos como GPT-4, Gemini, Claude 3, LLaMA 3, entre otros, han demostrado una capacidad notable para abordar tareas diversas en visión, lenguaje, recomendación, salud y educación. Su adaptabilidad y alcance multimodal abren nuevas oportunidades en múltiples sectores, aunque requieren consideraciones particulares respecto a recursos, transparencia y control.
- El desarrollo y despliegue de modelos de inteligencia artificial plantea retos significativos, especialmente en términos de sesgos, explicabilidad, privacidad, seguridad y concentración de poder. El análisis realizado confirma que, si bien la IA puede generar grandes beneficios sociales y económicos, su uso indiscriminado o sin salvaguardas adecuadas puede amplificar desigualdades, comprometer la confianza social y dificultar la equidad y la protección de derechos fundamentales. Es imprescindible continuar impulsando la investigación en IA ética, auditabilidad y regulación responsable.
- Las tendencias actuales muestran rápida evolución, orientada hacia modelos cada vez más multimodales, eficientes, explicables y democratizados. La personalización, la colaboración humano-IA y el surgimiento de marcos regulatorios robustos marcarán el desarrollo de la inteligencia artificial en los próximos años. Estas tendencias no solo abrirán nuevas oportunidades de innovación, sino que también requerirán una adaptación proactiva de los sistemas educativos, productivos y sociales para maximizar el beneficio colectivo y mitigar los riesgos inherentes.

Referencias

AI, Meta. 2024. «Llama 3: Open Foundation and Instruction-Tuned Models». https://ai.meta.com/llama/.

AI, Mistral. 2023. «Mixtral 8x7B». https://mistral.ai/news/mixtral-of-experts/.

Anthropic. 2024. «Introducing Claude 3». https://www.anthropic.com/news/claude-3-family.

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. Springer.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, y et al. 2021. «On the Opportunities and Risks of Foundation Models». *arXiv preprint arXiv:2108.07258*. https://arxiv.org/abs/2108.07258.

Cohere. 2024. «Introducing Command R: Next-Generation RAG Model». https://docs.cohere.com/docs/command-r.

DeepMind, Google, OpenAI, et al. 2023. «Gemini: A Family of Highly Capable Multimodal Models». *arXiv preprint arXiv:2312.11805*. https://arxiv.org/abs/2312.11805.

Goodfellow, Ian, Yoshua Bengio, y Aaron Courville. 2016. *Deep Learning*. MIT Press.

IBM. 2024. «¿Qué es un modelo de IA?» 2024. https://www.ibm.com/es-es/think/topics/ai-model.

Kirillov, Alexander, Eric Mintun, Nikhila Ravi, y et al. 2023. «Segment Anything». *arXiv preprint arXiv:2304.02643*. https://arxiv.org/abs/2304.02643.

OpenAI. 2023. «GPT-4 Technical Report». https://cdn.openai.com/papers/gpt-4.pdf.

———. 2024. «Introducing GPT-4o». https://openai.com/index/hello-gpt-4o.

Radford, Alec, Jong Wook Kim, Chris Hallacy, y et al. 2021. «Learning Transferable Visual Models From Natural Language Supervision». *arXiv preprint arXiv:2103.00020*. https://arxiv.org/abs/2103.00020.

Rouhiainen, Lasse. 2018. *Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro*. Barcelona: Alienta Editorial.

Russell, Stuart J., y Peter Norvig. 2010. *Inteligencia artificial: Un enfoque moderno*. Madrid: Pearson Educación.

xAI. 2023. «Grok: Open Source Models». https://x.ai/blog/grok-open-release.