🧠 Introducción a la Inteligencia Artificial

# ¿Qué es la Inteligencia Artificial?

La Inteligencia Artificial (IA) es una rama de la informática que busca crear sistemas capaces de realizar tareas que, si fueran hechas por humanos, requerirían inteligencia. Entre estas tareas se incluyen el reconocimiento de voz, visión por computadora, traducción automática, razonamiento lógico, juegos estratégicos, y más.

Existen distintos tipos de IA, desde sistemas con reglas simples hasta modelos avanzados que aprenden automáticamente a partir de datos.

# ¿Cómo aprende una IA?

El aprendizaje automático o Machine Learning es el proceso por el cual una IA mejora su rendimiento a partir de la experiencia. En lugar de ser programada con reglas fijas, la IA es entrenada con grandes cantidades de datos.

El proceso básico incluye:

1. Recolectar datos de entrenamiento (imágenes, textos, sonidos...).

2. Procesar y etiquetar esos datos.

3. Elegir un modelo de IA (por ejemplo, una red neuronal).

4. Entrenar el modelo ajustando sus parámetros internos hasta que aprenda a responder correctamente.

5. Validar el modelo y probarlo con datos nuevos.

# ¿Qué son los modelos de lenguaje?

Un modelo de lenguaje es un tipo de IA entrenada para entender, generar y continuar texto de forma coherente. Estos modelos se entrenan con enormes volúmenes de texto extraído de libros, sitios web, artículos y otros medios.

Modelos populares incluyen GPT (Generative Pre-trained Transformer), BERT, y LLaMA. Cada uno tiene arquitecturas diferentes, pero todos comparten la idea de aprender patrones en el lenguaje para generar respuestas relevantes.

# ¿Qué es un token y cómo se usa en IA?

En procesamiento de lenguaje natural, un 'token' es una unidad básica de texto: puede ser una palabra, parte de una palabra, o un símbolo. Por ejemplo, 'perro', ',', y 'inteligencia' pueden ser tokens.

Los modelos de lenguaje no procesan texto como frases completas, sino como secuencias de tokens. El modelo analiza el contexto de esos tokens para predecir qué palabra (token) viene después.

# ¿Qué es el contexto y por qué es importante?

El contexto es la información previa que se le da al modelo para que entienda mejor qué respuesta debe generar. Mientras más amplio el contexto, mejor puede el modelo comprender de qué se está hablando.

Por ejemplo, si decimos 'María compró una pelota. Luego ella...', el modelo usará el contexto para saber que 'ella' se refiere a María.

# ¿Qué es un prompt?

Un prompt es el texto de entrada que se le da a una IA para que genere una respuesta. Es la 'pregunta' o 'instrucción'.

Ejemplos:

-'¿Qué es la energía solar?'

-'Escribí una historia corta sobre un robot perdido.'

# ¿Qué tipos de modelos de IA existen?

1. IA simbólica: basada en reglas lógicas (si ocurre A, entonces hacer B).

2. IA estadística: basada en datos y probabilidades.

3. IA conexionista (redes neuronales): basada en neuronas artificiales que simulan el funcionamiento del cerebro.

4. Modelos de Deep Learning: redes neuronales profundas con múltiples capas, como los modelos actuales de IA generativa.

# ¿Cómo interactuamos con una IA?

Los usuarios interactúan con modelos de IA a través de aplicaciones, interfaces web o APIs. Al escribir un mensaje (prompt), el modelo analiza el texto, evalúa el contexto y genera una respuesta con base en su entrenamiento.

# Resumen visual del proceso de IA generativa

1. El usuario escribe un prompt.

2. El sistema lo convierte en tokens.

3. El modelo analiza el contexto y genera tokens de salida.

4. Los tokens se ensamblan como texto final.

5. La respuesta es mostrada al usuario.

# Conclusión

La Inteligencia Artificial, especialmente los modelos de lenguaje, representa una de las tecnologías más disruptivas de nuestro tiempo. Entender cómo funciona desde cero permite no solo usarla de forma efectiva, sino también formar parte activa de su evolución.

## IA débil vs IA fuerte

La IA débil (o IA estrecha) está diseñada para realizar una tarea específica, como reconocer rostros o sugerir películas. Es el tipo más común y está presente en asistentes virtuales como Siri o Alexa. La IA fuerte, por otro lado, busca igualar o superar la inteligencia humana en múltiples tareas, incluyendo el razonamiento, la creatividad y la comprensión del lenguaje, pero todavía es una meta lejana.

## Tipos de aprendizaje en IA

1. Aprendizaje supervisado: el modelo aprende a partir de ejemplos etiquetados (por ejemplo, imágenes con la etiqueta 'gato' o 'perro').

2. Aprendizaje no supervisado: el modelo explora datos sin etiquetas, buscando patrones o agrupamientos naturales.

3. Aprendizaje por refuerzo: el modelo aprende a través de prueba y error, recibiendo recompensas o penalizaciones (como en los videojuegos o en la robótica).

## ¿Qué es un transformer?

Los transformers son una arquitectura de red neuronal revolucionaria que permite a los modelos procesar palabras en contexto. En lugar de analizar el texto palabra por palabra, los transformers consideran todas las palabras al mismo tiempo (en paralelo), permitiendo generar respuestas más coherentes y profundas.

GPT, BERT y otros modelos modernos están basados en transformers. Esta arquitectura es la base del procesamiento moderno del lenguaje natural (PLN).

## Relación entre tokens y contexto

Un modelo de lenguaje tiene una 'ventana de contexto' que limita la cantidad de tokens que puede procesar al mismo tiempo. Por ejemplo, un modelo como GPT-3.5 puede manejar hasta 4.096 tokens en una sola entrada. Esto incluye tanto el prompt como la respuesta generada. Si se supera este límite, el modelo 'olvida' los tokens más antiguos.

## Ingeniería de prompts

La forma en que se formula el prompt afecta directamente la calidad de la respuesta. Es posible mejorar los resultados mediante técnicas como:

- Ejemplificación: agregar ejemplos claros de lo que se espera.

- Roles: pedirle al modelo que actúe como un especialista ('Actuá como un docente de IA...').

- Encadenamiento: construir prompts que incluyan pasos intermedios para tareas complejas.

## Modelos generativos vs discriminativos

Los modelos generativos, como GPT, crean texto desde cero. Son útiles para redacción, síntesis, escritura creativa, etc. Los modelos discriminativos, en cambio, se enfocan en clasificar datos o predecir etiquetas. Se usan en tareas como detección de spam, reconocimiento facial o diagnóstico médico.

## APIs y entornos de uso

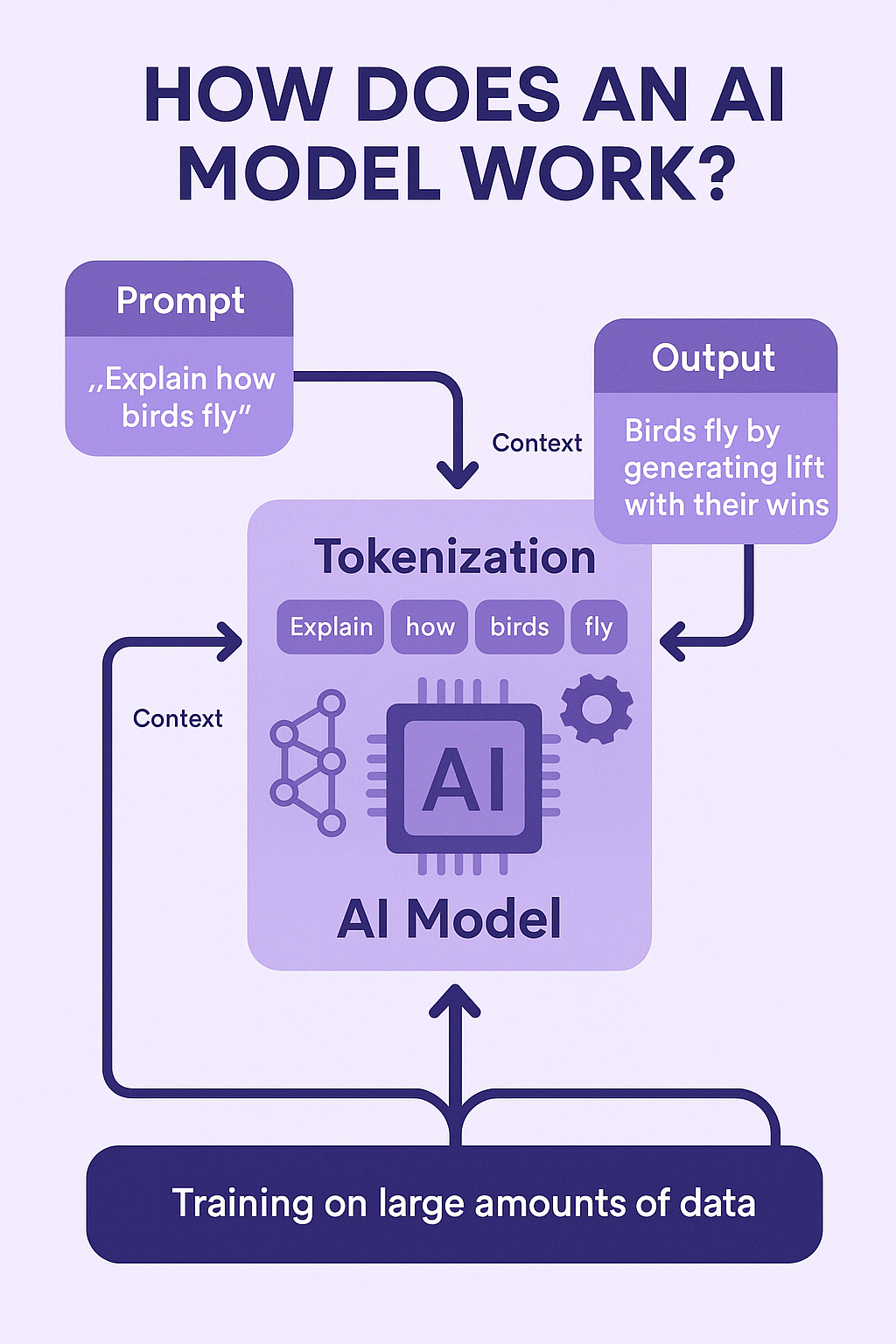
Los modelos de lenguaje pueden integrarse en aplicaciones a través de APIs como OpenAI, HuggingFace o Cohere. Estas interfaces permiten enviar texto al modelo y recibir respuestas, facilitando su integración en páginas web, asistentes virtuales, bots, sistemas de soporte, entre otros.

## Capacidades y límites de una IA actual

✔️ Puede: resumir textos, generar ideas, traducir idiomas, escribir código, responder preguntas, etc.

❌ No puede: razonar como un humano, tener conciencia, acceder a emociones, tener intenciones.

Es fundamental entender que la IA no 'piensa' como una persona, sino que responde en función de patrones estadísticos aprendidos.



# 🖼️ Explicación de la infografía: ¿Cómo funciona un modelo de IA?

La infografía presentada ilustra el funcionamiento interno de un modelo de Inteligencia Artificial basado en lenguaje, como los modelos GPT (Generative Pre-trained Transformer). Este proceso puede dividirse en varias etapas clave:  
  
1. \*\*Entrada del usuario (Prompt)\*\*: Todo comienza con el texto que el usuario introduce. Este puede ser una pregunta, una instrucción, una solicitud creativa, etc.  
  
2. \*\*Tokenización\*\*: El texto es dividido en unidades más pequeñas llamadas tokens, que son las piezas mínimas de información que el modelo puede procesar (pueden ser palabras, sílabas o caracteres según el modelo).  
  
3. \*\*Análisis del contexto\*\*: Los tokens son procesados dentro de una ventana de contexto. El modelo evalúa la relación entre todos los tokens dentro de esa ventana para entender la intención y el significado del mensaje.  
  
4. \*\*Capa de atención\*\*: Esta es una de las partes clave de los transformers. Permite que el modelo enfoque su atención en las palabras más relevantes del texto para generar una respuesta coherente.  
  
5. \*\*Predicción del siguiente token\*\*: Basado en lo anterior, el modelo predice cuál es el token más probable que debería venir a continuación. Este proceso se repite iterativamente para generar una secuencia completa de salida.  
  
6. \*\*Conversión de tokens a texto\*\*: Una vez generados los tokens de salida, se convierten nuevamente en texto legible para el usuario.  
  
7. \*\*Respuesta del modelo\*\*: El resultado final es presentado como una respuesta fluida y contextual, visible para el usuario.  
  
Este flujo permite entender cómo una IA transforma una entrada textual en una respuesta compleja usando redes neuronales de múltiples capas y procesamiento de lenguaje natural a gran escala.