

CAPÍTULO V

IA GENERATIVA Y MODELOS DE LENGUAJE

OBJETIVO:

El objetivo de este capítulo es introducir al lector en el funcionamiento, las capacidades y el ecosistema global de las inteligencias artificiales generativas, con especial énfasis en los modelos de lenguaje de gran escala (LLMs) y su impacto en el ámbito educativo, científico y tecnológico.

Se analiza cómo operan estos modelos, qué tecnologías los sustentan, qué actores lideran su desarrollo a nivel mundial, y cuáles son sus aplicaciones más destacadas. Además, se presentan comparaciones entre las principales plataformas disponibles (como ChatGPT, Gemini o Claude), se abordan conceptos técnicos fundamentales —como tokens, ventanas de contexto y APIs— y se describe la dinámica geopolítica de la carrera por la supremacía en inteligencia artificial entre potencias como Estados Unidos y China.

Este capítulo proporciona las bases para comprender las capacidades, limitaciones y usos estratégicos de las IA generativas en la educación y otros sectores clave del conocimiento y la innovación.

TEMAS A EXPLORAR:

- Qué es la inteligencia artificial generativa, cómo funciona y por qué representa uno de los avances más disruptivos de la era digital.
- Cuáles son las fases clave del funcionamiento de estas IA (entrenamiento masivo, predicción secuencial y afinamiento), y qué ventajas y limitaciones presentan.
- Cómo se aplican herramientas como ChatGPT en educación, investigación y creatividad, y qué principios tecnológicos las sustentan (Transformers, atención, tokens, contextualización).
- Qué tipos de contenido puede generar una IA: texto, imágenes, audio, video y código, y cómo esto impacta diferentes industrias.
- Qué es un modelo de lenguaje de gran escala (LLM), cómo se entrena y qué lo hace diferente del pensamiento humano.
- Qué rol juegan los tokens y las ventanas de contexto en la calidad y coherencia de las respuestas de una IA.

- Qué son las APIs y cómo permiten llevar la IA a aplicaciones reales: educación personalizada, automatización de tareas, asistentes virtuales y más.
- Cuáles son las diferencias entre modelos de código abierto y cerrado, y cómo afecta esto a la ética, seguridad y soberanía tecnológica.
- Quiénes lideran la carrera global por el desarrollo de LLMs, y cómo se configuran los ecosistemas de IA en Estados Unidos, China y otros actores internacionales.

CHATGPT EN ACCIÓN

- ¿Qué puede hacer ChatGPT hoy por ti como educador, creador o estudiante?
- Tokens en la práctica: ¿cuántos puedes usar sin que se pierda el hilo?
- Dilemas éticos: ¿cómo evitar la desinformación o el uso irresponsable de la IA?
- Comparando plataformas: ¿cuál IA se adapta mejor a tu contexto?
- Resumen activo: ¿Qué aprendiste sobre LLM, plataformas y el futuro de ChatGPT?

HERRAMIENTAS IA: Elicit

5.1. ¿Qué es la IA generativa? Fundamentos y aplicaciones

La inteligencia artificial generativa representan uno de los avances más disruptivos de la era digital. A diferencia de otras formas de inteligencia artificial que simplemente clasifican datos o responden de forma preestablecida, estas tecnologías tienen la capacidad de crear contenido nuevo —ya sea texto, imágenes, código, audio o video— a partir de patrones aprendidos.

Pero, ¿cómo es posible que una máquina escriba un poema, resuma un artículo académico o genere una imagen de una ciudad que no existe? Para comprenderlo, es necesario explorar los fundamentos de su funcionamiento.

5.1.1. ¿Qué es la IA generativa?

Las IA generativas son modelos computacionales entrenados con cantidades masivas de datos, diseñados para predecir y generar información coherente y contextual. Son capaces de “imaginar” contenido nuevo basándose en lo que han aprendido durante su fase de entrenamiento. El ejemplo más conocido es ChatGPT, que genera texto conversacional a partir de una simple entrada del usuario.

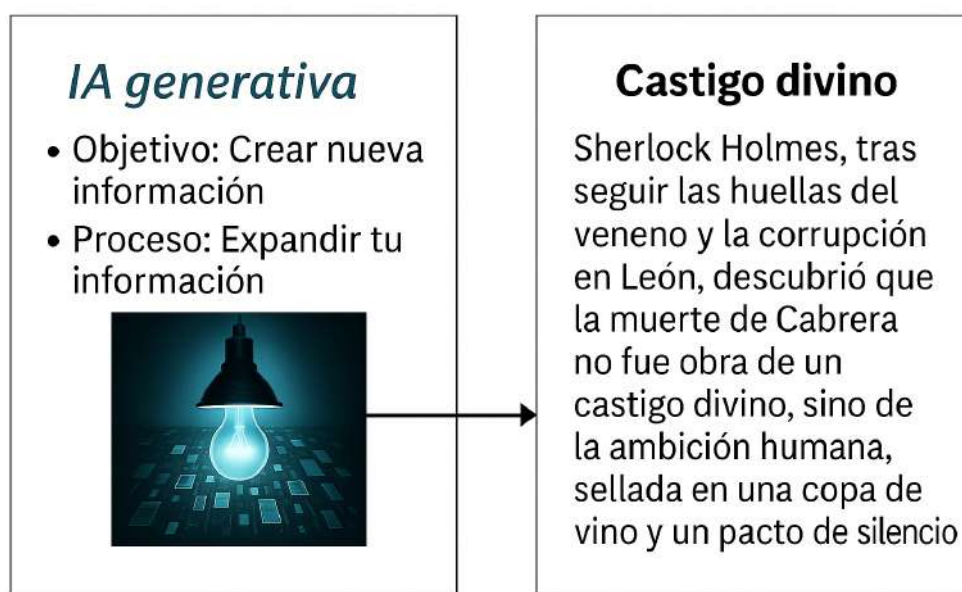


Figura 5.1. IA Generativa en acción. Crear una historia corta de Sherlock Holmes resolviendo el caso de Castigo Divino de Sergio Ramírez

¿Cómo funcionan?

El proceso puede entenderse en tres fases clave:

1. Entrenamiento masivo

Los modelos son expuestos a billones de palabras, imágenes, líneas de código o notas musicales. Este entrenamiento no es manual, sino que se basa en técnicas de aprendizaje profundo, particularmente en redes neuronales transformadoras (Transformers), que identifican patrones, relaciones y estructuras complejas en los datos.

ADQUISICIÓN DE LENGUAJE

Maximizar la exposición a materiales



Figura 5.2. Adquisición de Lenguaje de los Modelos

2. Predicción secuencial

En el caso de los modelos de texto, como ChatGPT, el sistema aprende a predecir la siguiente palabra en una oración con base en las anteriores. Esta simple lógica —repetida millones de veces— permite construir oraciones completas, párrafos, respuestas y hasta ensayos enteros que parecen escritos por humanos.

IA generativa

- Objetivo: Crear nueva información
- Proceso: Expandir tu información

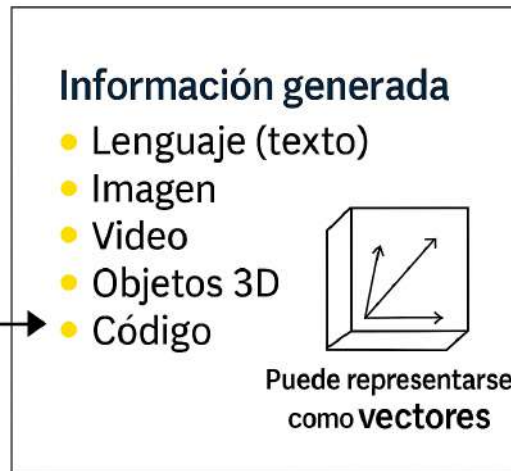


Figura 5.3. Conversión en Vectores

3. Afinamiento (fine-tuning)

Una vez entrenado, el modelo puede ser afinado con datos específicos o supervisión humana para realizar tareas más precisas: redactar correos, responder preguntas, resolver problemas o conversar con fluidez y coherencia.

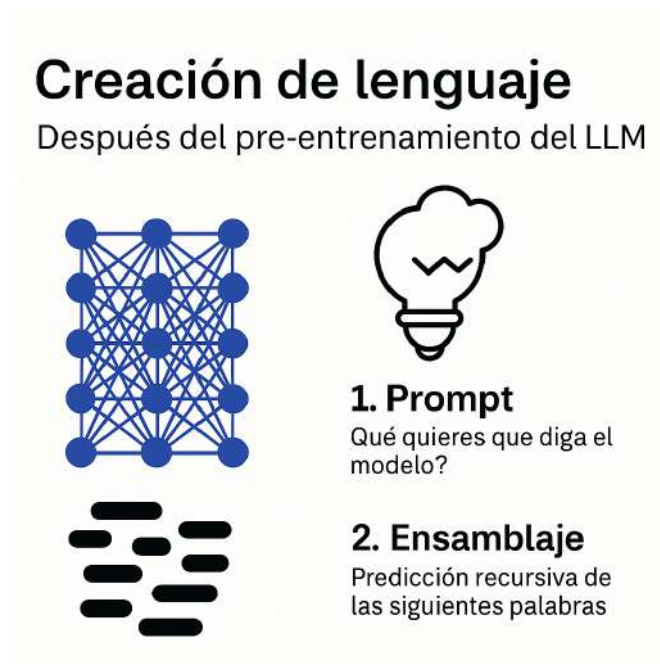


Figura 5.4. Afinamiento de los modelos

Ventajas

- Automatización de tareas cognitivas: Las IA generativas pueden redactar, resumir, traducir, codificar o diseñar, liberando tiempo y recursos humanos.
- Asistencia en procesos creativos: Ofrecen borradores de textos, ideas visuales o simulaciones que sirven de punto de partida para creadores, docentes o investigadores.
- Escalabilidad y velocidad: Generan contenido en segundos, permitiendo implementar soluciones educativas, empresariales o sociales a gran escala.

Limitaciones y desafíos

- Sesgos y errores: Los modelos reflejan los sesgos presentes en los datos con los que fueron entrenados. Esto puede llevar a generar respuestas incorrectas, parciales o culturalmente sesgadas.
- Falta de comprensión real: Aunque el modelo genera lenguaje natural, no “entiende” el contenido como lo haría un ser humano; simplemente reproduce patrones lingüísticos.
- Riesgos éticos: Su mal uso puede facilitar la creación de desinformación, manipulación o sustitución indebida del pensamiento crítico.



Figura 5.5. Entornos Cognitivos de Alta Tecnología

Podemos resumir los puntos anteriores en la siguiente tabla:

Tabla 5.1. Características y Funcionamiento de las IA Generativas

Elemento	Descripción
Qué son	Modelos capaces de generar contenido nuevo (texto, imagen, audio, código).
Cómo funcionan	Usan redes neuronales para predecir y generar contenido secuencial.
Ventajas	Automatización, creatividad, rapidez.
Limitaciones	Riesgo de sesgos, errores y mal uso.

Las IA generativas no son simples herramientas de automatización: representan un nuevo paradigma cognitivo artificial. Comprender su lógica de funcionamiento es el primer paso para utilizarlas de forma crítica, ética y transformadora, especialmente en contextos educativos.

5.1.2. ¿Qué hace ChatGPT y otras herramientas generativas?

Uno de los ejemplos más representativos de inteligencia artificial generativa es ChatGPT, desarrollado por la organización OpenAI. Este modelo ha transformado la manera en que interactuamos con las máquinas al ofrecer respuestas coherentes, creativas y contextuales en lenguaje natural.

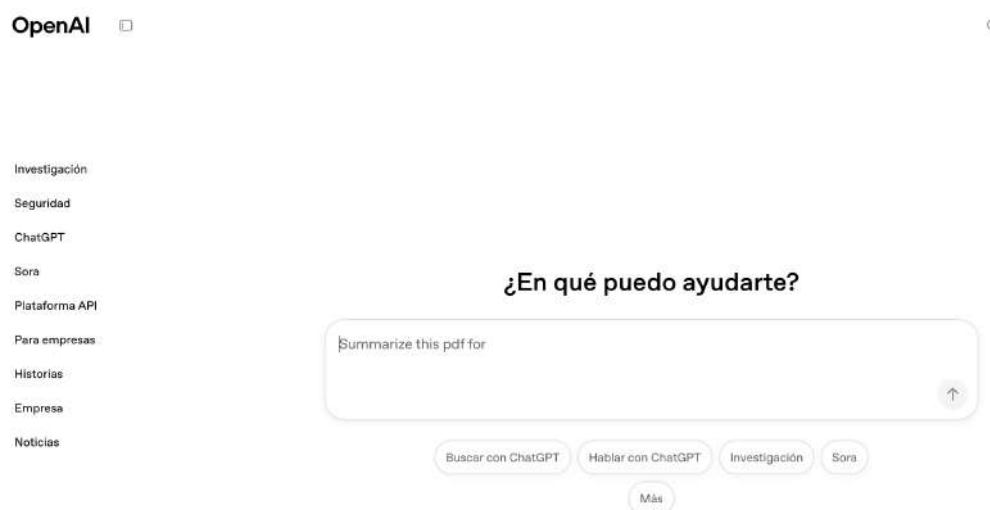


Figura 5.6 Página principal de [OpenAI.com](https://openai.com), específicamente del **asistente de ChatGPT** en su interfaz web. <https://openai.com>

Aunque su funcionamiento puede parecer “mágico” desde fuera, en el fondo ChatGPT se basa en principios matemáticos y estadísticos sofisticados que lo hacen sorprendentemente eficaz para tareas comunicativas y cognitivas complejas.

Predicción de la siguiente palabra

En esencia, ChatGPT funciona prediciendo la siguiente palabra más probable en una secuencia de texto. Para lograr esto, fue entrenado con una cantidad masiva de información textual —libros, artículos, sitios web, conversaciones— y aprendió a identificar patrones lingüísticos que le permiten construir frases con sentido.

Este mecanismo de predicción palabra por palabra no es trivial: requiere comprender el contexto, la estructura gramatical, el estilo del interlocutor y la intención del mensaje.



Figura 5.7 Modelos de Lenguaje en Acción

Por ejemplo, ante la frase incompleta:

“La educación es la clave para...”

ChatGPT puede completar con múltiples opciones posibles, como “el desarrollo personal”, “una sociedad más justa” o “el progreso humano”, según el tono, el tema y la conversación previa.

Comprensión del contexto y relaciones entre palabras

A diferencia de los modelos anteriores, ChatGPT no analiza las palabras de forma aislada, sino que las relaciona entre sí en contextos complejos. Gracias a la arquitectura Transformer y el mecanismo de atención, puede recordar e interpretar fragmentos largos de texto y mantener la coherencia a lo largo de una conversación.

Esta capacidad para conectar ideas y adaptar su estilo a diferentes registros —formal, coloquial, técnico o creativo— es lo que le permite escribir desde correos electrónicos hasta cuentos infantiles o resúmenes académicos.

Un salto evolutivo: “Attention Is All You Need”

El avance decisivo que permitió el desarrollo de modelos como ChatGPT fue el artículo “Attention Is All You Need”, publicado en 2017 por investigadores de Google. Este trabajo presentó una nueva arquitectura llamada Transformer, que revolucionó el procesamiento del lenguaje natural (NLP).

En lugar de procesar una frase de principio a fin, los Transformers “prestan atención” a todas las palabras al mismo tiempo, identificando cuáles son más relevantes para comprender o generar una respuesta adecuada. Esta innovación resolvió un problema crónico de los modelos anteriores: la incapacidad de recordar o conectar elementos distantes dentro de un texto.

Este principio es tan poderoso que, años después, ha sido citado decenas de miles de veces y ha dado origen a herramientas como ChatGPT, Gemini, Claude y muchas otras.

Aplicaciones derivadas

El impacto de esta tecnología no se limita al texto. Hoy, modelos basados en arquitecturas similares están siendo utilizados para crear música (como Suno), generar imágenes (como DALL·E o Midjourney), diseñar juegos o incluso automatizar procesos educativos completos.

Un ejemplo ilustrativo

Consideremos una pregunta humorística e inusual como:

"¿Por qué el koala eructó educadamente mientras recitaba pi hasta cincuenta decimales?"

Aunque absurda, ChatGPT es capaz de generar una respuesta lógica, creativa y con sentido dentro del contexto del lenguaje. ¿Cómo lo hace?

- 1. Analiza los componentes semánticos: identifica a un sujeto (koala), una acción peculiar (eructar educadamente), y una situación abstracta (recitar pi)
- 2. Usa el mecanismo de atención para relacionar estas ideas de manera coherente.
- 3. Genera una respuesta adaptada al tono humorístico de la pregunta.

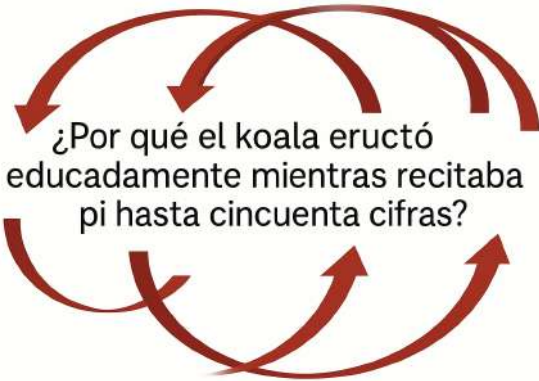


Figura 5.8 Visualización del Mecanismo de Atención

Una posible respuesta podría ser:

“El koala eructó educadamente para no interrumpir la atmósfera solemne de su recital matemático, mientras demostraba su increíble memoria numérica.”

Este tipo de respuesta no solo muestra la capacidad técnica del modelo, sino también su flexibilidad creativa y contextual, cualidades que lo hacen tan valioso en el entorno educativo.

En resumen, ChatGPT funciona de la siguiente manera:

Tabla 5.2.Funcionamiento del ChatGPT

Función	Descripción
Predicción de texto	Anticipa la siguiente palabra según el contexto.
Comprensión contextual	Relaciona ideas, mantiene coherencia y adapta su estilo.
Base tecnológica	Mecanismo de atención y la arquitectura Transformer.
Versatilidad de uso	Puede responder preguntas, redactar textos, generar ideas o traducir.
Creatividad controlada	Capaz de manejar desde temas serios hasta preguntas absurdas con coherencia.

ChatGPT no “entiende” el lenguaje como lo haría un ser humano, pero su capacidad de imitar patrones complejos lo convierte en una herramienta poderosa, versátil y sorprendentemente cercana al pensamiento humano, ideal para apoyar procesos de aprendizaje, investigación y creación.

5.1.3 Características de las IA generativas

Las inteligencias artificiales generativas, especialmente aquellas basadas en modelos de lenguaje de gran escala (LLMs), no solo destacan por su capacidad para generar contenido, sino también por una serie de características técnicas que las hacen extraordinariamente poderosas, adaptables y eficaces.

En esta sección exploraremos sus principales atributos funcionales, estructurales y lingüísticos, que explican por qué estas herramientas están transformando la educación, la investigación y la creatividad digital.

Tabla 5.3. Principales Características de la IA Generativa

Característica	Descripción breve
Arquitectura Transformer	Procesamiento paralelo del texto, permitiendo mayor velocidad, coherencia y comprensión contextual.
Aprendizaje dual	Fase de preentrenamiento masivo + fine-tuning específico.
Adaptabilidad	Puede realizar múltiples tareas sin reentrenamiento, como traducir, redactar, resumir o programar.
Contextualización	Analiza conversaciones previas para mantener coherencia en respuestas largas.
Multilingüismo	Responde en varios idiomas (humanos y de programación).
Creatividad algorítmica	Capaz de responder a preguntas abiertas o incluso absurdas de forma coherente.

1. Arquitectura Transformer

El núcleo tecnológico que impulsa a la IA generativa moderna es la arquitectura Transformer, introducida en 2017. A diferencia de las redes neuronales tradicionales, los Transformers procesan las secuencias de texto en paralelo en lugar de hacerlo palabra por palabra. Esto permite:

- Mayor velocidad y eficiencia computacional.
- Captura de relaciones semánticas a largo plazo entre palabras, incluso si están muy distantes en un texto.
- Respuestas más coherentes y naturales.

- El mecanismo de atención incorporado en esta arquitectura permite al modelo "enfocarse" en las partes relevantes del texto para generar una respuesta significativa, contextualizada y precisa.

2. Fase de preentrenamiento y afinamiento

Los modelos generativos pasan por dos etapas principales de desarrollo:

- Preentrenamiento: Durante esta fase, el modelo se expone a una enorme cantidad de texto (libros, artículos, conversaciones, código, etc.) y aprende los patrones estadísticos del lenguaje sin una tarea específica. Es como si "leyera" todo Internet para familiarizarse con cómo escriben los humanos.
- Afinamiento o fine-tuning: Una vez preentrenado, el modelo es ajustado con datos más específicos y con supervisión humana para orientarlo hacia tareas concretas, como redacción de correos, resolución de problemas, generación de resúmenes o asistencia pedagógica.

Este proceso dual permite que los modelos sean versátiles y altamente funcionales, adaptándose a diferentes contextos de uso.

3. Generación de texto coherente y creativo

Las IA generativas son capaces de crear contenido textual fluido, gramaticalmente correcto y estilísticamente adaptado. Pueden redactar en tono académico, informal, técnico o literario, y generar:

- Resúmenes y explicaciones.
- Artículos, ensayos o informes.
- Diálogos naturales para asistentes virtuales.
- Contenido educativo y actividades interactivas

Además, su creatividad algorítmica permite responder incluso a preguntas abiertas o absurdas con una sorprendente coherencia.

4. Adaptabilidad y diversidad de tareas

Una de las fortalezas clave de estos modelos es su capacidad de adaptación a distintos tipos de tareas:

- Traducción automática multilingüe.
- Generación de contenidos educativos personalizados.
- Respuesta a preguntas complejas o abstractas.
- Soporte en escritura creativa, técnica o publicitaria.

Todo esto se logra sin necesidad de reentrenar el modelo desde cero, lo que hace su implementación más accesible para instituciones educativas, empresas y desarrolladores.

5. Contextualización de las respuestas

Los modelos generativos como ChatGPT utilizan su ventana de contexto para comprender lo que el usuario ha dicho anteriormente y generar respuestas alineadas con esa información. Esta habilidad permite mantener la coherencia en conversaciones largas, resolver dudas progresivamente o ajustar el tono según el interlocutor.

Por ejemplo, si en una conversación previa el usuario mencionó que es docente universitario, el modelo puede adaptar sus respuestas a ese rol, proporcionando ejemplos, sugerencias o aplicaciones prácticas relacionadas con la docencia.

6. Manejo de múltiples idiomas

Aunque la mayoría de los modelos han sido entrenados principalmente en inglés, la exposición multilingüe a grandes cantidades de texto ha permitido que puedan responder con soltura en una variedad de idiomas. Esto incluye:

- Idiomas romances como español, francés o italiano.
- Lenguas germánicas como alemán o neerlandés.
- Lenguajes de programación como Python, HTML o SQL.

Si bien el rendimiento no es igual en todos los idiomas, la capacidad para entender y producir texto en diferentes lenguas abre un abanico de posibilidades para contextos educativos multilingües.

5.1.4 Tipos de contenido que puede generar

La inteligencia artificial generativa no es solo una herramienta; es una fábrica creativa que nunca duerme. Una mente digital capaz de transformar ideas abstractas en productos concretos, en cuestión de segundos. Su poder no se limita a escribir textos o generar imágenes curiosas. Estamos hablando de una tecnología que está reconfigurando el proceso mismo de creación en múltiples industrias: educación, publicidad, entretenimiento, salud, desarrollo de software y más.

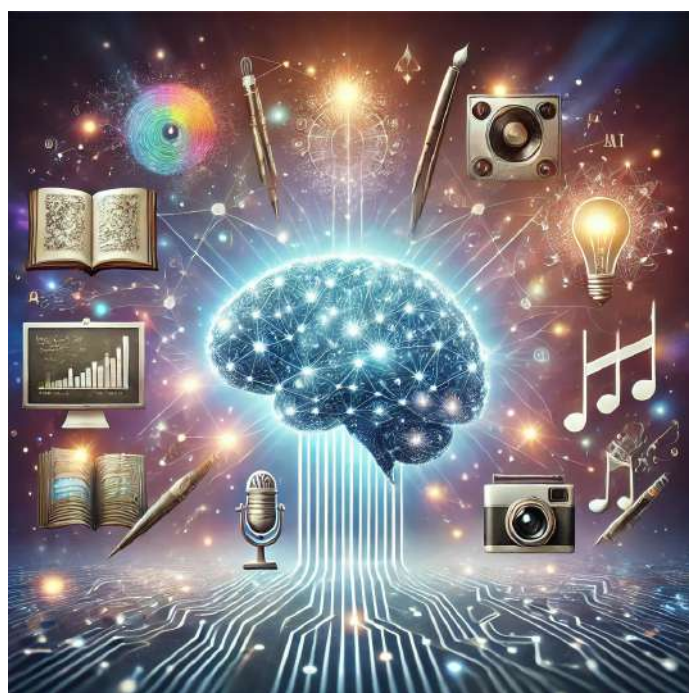


Figura 5.9 La Mente Multimodal de la IA

Para comprender realmente el alcance de esta revolución, es fundamental conocer los principales tipos de contenido que una IA puede generar. Porque cuanto más amplio sea nuestro conocimiento de sus capacidades, mayor será nuestra posibilidad de aprovecharlas estratégicamente.

1. TEXTO: de la idea al discurso en segundos

La generación de texto ha sido una de las puertas de entrada más populares al mundo de la IA. Y no es casualidad: el lenguaje es nuestra herramienta básica de comunicación, enseñanza, persuasión y creación.

Las IA actuales pueden redactar contenido de una calidad que, en muchos casos, rivaliza (y hasta supera) lo que un ser humano promedio podría lograr en tiempo limitado.

¿Qué puede generar en formato texto?

- Artículos periodísticos, ensayos argumentativos, columnas de opinión
- Posts para redes sociales, titulares publicitarios y copies persuasivos
- Emails profesionales, cartas de presentación y discursos públicos
- Guiones para videos de YouTube, TikToks o presentaciones corporativas
- Cuentos, poemas, diálogos teatrales o incluso novelas enteras
- Resúmenes de textos largos, traducciones, reescritura de estilo o tono
- Respuestas a preguntas académicas, explicaciones paso a paso, ejercicios

¿Por qué es tan poderoso esto?

Porque permite escalar la comunicación escrita sin necesidad de un ejército de redactores. También democratiza la posibilidad de expresarse profesionalmente, incluso para quienes no dominan ciertos lenguajes o estilos. Además, sirve como punto de partida: una “primera versión” rápida que se puede editar, afinar y personalizar.

2. IMAGEN: pintar con palabras

Con solo describir una escena, una emoción o una idea, las IA pueden generar imágenes realistas, conceptuales o artísticas. Esto ha revolucionado el mundo del diseño, la ilustración y la publicidad.

¿Qué tipos de imágenes puede crear una IA?

- Ilustraciones para libros, portadas de eBooks, o materiales educativos
- Diseño gráfico para campañas publicitarias, logos, infografías
- Arte conceptual para videojuegos, películas o marcas personales
- Estilo fotográfico (retratos, paisajes, arquitectura) para contenido visual
- Prototipos de productos, decoración de interiores, moda
- Personajes originales para cómics, animaciones o proyectos creativos

Imagina esto: un emprendedor puede visualizar la estética de su marca antes de pagarle a un diseñador. Un profesor puede ilustrar conceptos abstractos para sus clases en minutos. Un autor puede crear portadas impactantes sin tener conocimientos técnicos de diseño.

La IA no elimina al artista, lo potencia.

3. AUDIO: darle voz, ritmo y atmósfera a tus ideas

El sonido también forma parte de esta nueva frontera creativa. Las IA generativas de audio permiten producir música, voces sintéticas con emociones, efectos de sonido e incluso narraciones completas.

¿Qué contenido auditivo puede generar una IA?

- Música original en distintos géneros (lo-fi, clásica, electrónica, ambiental)
- Voz sintética para audiolibros, tutoriales, chatbots, asistentes virtuales
- Narración de cuentos o textos largos con diferentes estilos de entonación
- Doblajes automáticos de videos en múltiples idiomas
- Efectos sonoros para videojuegos, películas, animaciones o apps
- Podcasts automatizados: voz, guion y música en un solo flujo creativo

¿Por qué es relevante esto?

Porque permite producir contenido accesible (por ejemplo, para personas con dificultades visuales), crear experiencias inmersivas en plataformas digitales, y generar materiales de aprendizaje más atractivos. Además, reduce los costos y tiempos que antes implicaban largas horas de grabación o edición.

4. VIDEO: de la palabra a la acción

Aunque está en una etapa más reciente, la generación de video por IA está avanzando a pasos de gigante. Lo que hace poco parecía ciencia ficción —crear videos completos con solo texto— hoy ya es una realidad.

¿Qué tipo de contenido en video puede generar una IA?

- Clips breves explicativos o promocionales para redes sociales
- Avatares animados que leen guiones o presentan información
- Doblaje automático con sincronización labial
- Videos educativos con gráficos dinámicos y voz sintética
- Conversión de textos a videos visualmente atractivos (text-to-video)
- Animaciones simples para contar historias o representar conceptos

¿Y esto para qué sirve?

Para escalar la producción de contenido visual sin necesidad de cámaras, locaciones ni actores. También abre una puerta enorme para la educación a distancia, el marketing digital, la producción audiovisual de bajo presupuesto o el contenido personalizado.

5. CÓDIGO: programar sin ser programador

Finalmente, una de las capacidades más impresionantes de la IA: generar código. Esto no solo implica escribir líneas de programación, sino entender lo que el usuario necesita, proponer soluciones y explicarlas.

¿Qué tipo de código puede generar una IA?

- Scripts automatizados para tareas repetitivas (ej. Excel, Python, Bash)
- Aplicaciones web simples o prototipos funcionales
- Juegos, bots, extensiones de navegador o asistentes digitales
- Soluciones a problemas de programación y debugging de errores
- Traducción entre lenguajes de programación (de Python a JavaScript, por ejemplo)
- Comentarios explicativos y documentación técnica

Esto no solo ayuda a los expertos, también empodera a quienes están aprendiendo. La IA puede convertirse en una especie de mentor personal, acelerando el proceso de formación y reduciendo barreras técnicas. Y para los desarrolladores experimentados, es una manera de ahorrar tiempo, depurar ideas y mantenerse enfocados en lo estratégico.

Un universo creativo en expansión

Como puedes ver, las posibilidades no se limitan a un solo tipo de contenido. La IA es una herramienta transversal, un catalizador creativo, una extensión de nuestras capacidades humanas. No reemplaza la imaginación, pero sí la acelera. No sustituye el talento, lo amplifica.

Cada día surgen nuevas plataformas, nuevas funciones y nuevos modelos que hacen esta tecnología más precisa, accesible y poderosa. Entender los distintos formatos que puede generar es el primer paso para integrarla de forma estratégica en nuestras profesiones, proyectos o emprendimientos.

Porque en esta nueva era, el límite ya no es lo que sabes hacer tú solo... sino lo que puedes crear con tu copiloto inteligente.

5.2. LLM: El motor detrás de la IA generativa de texto

- ¿Qué es un LLM (Large Language Model)?
- Tokens: la unidad del lenguaje computacional
- Ventanas de contexto (Context Window)
- APIs y su rol en la integración de IA
- ¿Por qué algunos modelos son cerrados y otros de código abierto?

Si la inteligencia artificial generativa es el vehículo que nos está llevando a una nueva era de creación, los LLMs (Large Language Models) son el motor que la impulsa. Son el corazón latente detrás de cada texto que se genera, cada respuesta que leemos, cada correo redactado en segundos. Pero ¿qué son realmente? ¿Cómo funcionan? ¿Por qué algunos parecen “saber tanto”? ¿Y por qué unos son abiertos y otros no?



Figura 5.10 La Máquina que Aprende

En esta sección vamos a levantar el capó de esta tecnología. No hace falta ser programador ni científico de datos para entenderlo. Solo necesitas curiosidad y una pregunta en mente: *¿cómo es posible que una máquina escriba tan bien?*

5.2.1 ¿Qué es un LLM (Large Language Model)?

Si tuvieras que construir una inteligencia capaz de conversar contigo, responder preguntas, escribir historias, ayudarte con tareas complejas y hasta ofrecerte consejos, ¿por dónde empezarías?

La respuesta de la comunidad científica y tecnológica ha sido esta: entrenando modelos matemáticos con el lenguaje humano. Así nacen los LLMs: Large Language Models, o en español, Modelos de Lenguaje de Gran Escala.

🧠 Un lector incansable que aprende patrones

Imagina un lector que nunca duerme. Un lector que ha devorado millones de libros, artículos, ensayos, correos electrónicos, líneas de código, guiones de películas, foros de internet, enciclopedias y conversaciones digitales.

Ese lector no solo memoriza, sino que reconoce patrones: cómo se organiza una historia, cómo se argumenta una idea, cómo se estructura una respuesta a una pregunta común. Aprende qué palabras suelen venir después de otras, qué expresiones son propias de cierto contexto, qué significa un silencio entre líneas.

Ese lector, ese modelo entrenado para detectar y reproducir el lenguaje, es un LLM.

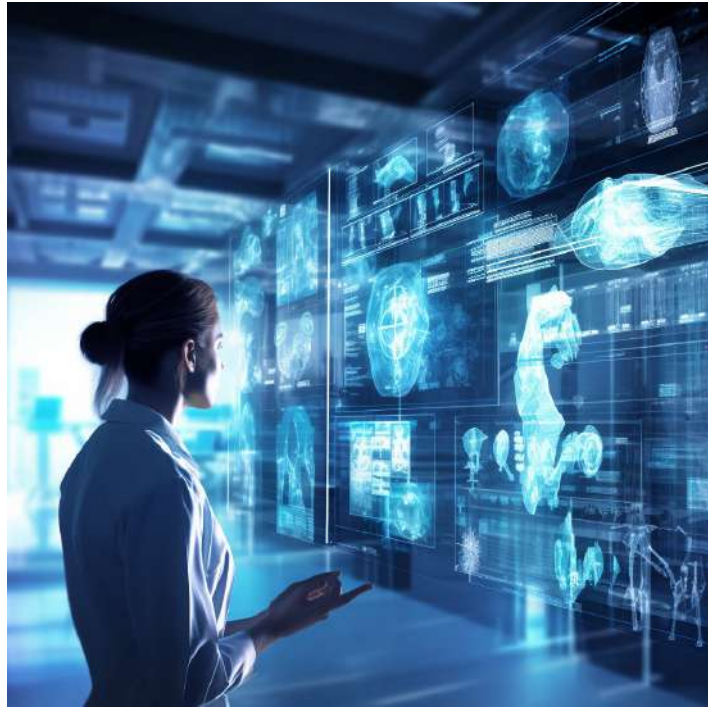


Figura 5.11. Visualización Médica Avanzada con IA

¿Cómo se entrena un LLM?

Un LLM no es una caja mágica. Es una estructura matemática extremadamente compleja, formada por lo que se conoce como red neuronal profunda. Para que este sistema funcione, necesita exponerse a cantidades colosales de datos textuales.

El proceso general es el siguiente:

1. Se recolectan datos: textos públicos de libros, sitios web, artículos científicos, redes sociales, Wikipedia, diálogos, código fuente, etc.
2. Se preprocesa la información: se convierte el lenguaje en tokens —pequeñas unidades de texto comprensibles para la máquina.
3. Se entrena el modelo: la red neuronal comienza a predecir el siguiente token en una secuencia, basándose en millones de ejemplos.
4. Se ajusta con retroalimentación: se evalúa qué tan cerca estuvo la predicción de la realidad y se corrige. Este proceso ocurre miles de millones de veces.

Es como si le dijeras a una IA:

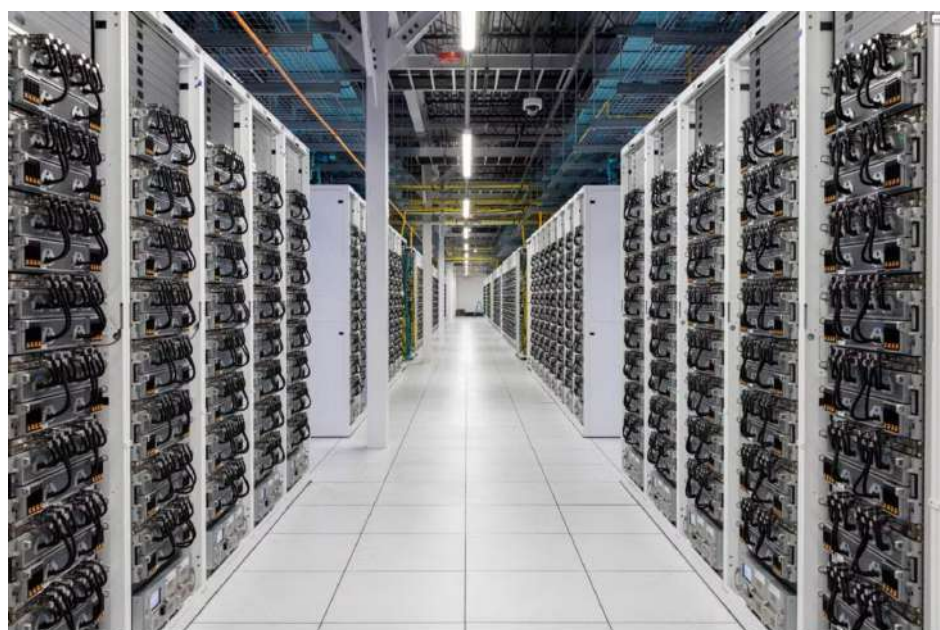
"Voy a darte una frase incompleta miles de millones de veces... y quiero que intentes adivinar cómo sigue."

Así, poco a poco, el modelo aprende a escribir, responder, sugerir, traducir, resumir y razonar en lenguaje natural.

¿Piensa un LLM como un humano?

La respuesta corta es: no.

La respuesta más precisa es: piensa de forma diferente.



Colossus Data Center Compute Hall

Figura 5.12 Colossus xAI – Memphis. Centro de datos de Elon Musk con más de 200,000 GPUs. Construido en 122 días. Meta: alcanzar 1 millón de GPUs y liderar en IA global.

Un LLM no tiene emociones, creencias, intenciones ni conciencia. No comprende el mundo como tú o yo, ni tiene una experiencia subjetiva. Pero lo que sí tiene es una increíble capacidad para reconocer cómo usamos el lenguaje y para predecir cómo lo usaríamos a continuación.

- ♦ No sabe qué es la tristeza, pero puede escribir un poema triste.
- ♦ No entiende la justicia, pero puede explicarte teorías jurídicas.
- ♦ No ha vivido una experiencia humana, pero puede describirla con fidelidad.

Esto se debe a que ha visto cómo millones de humanos hablan, escriben, sienten, argumentan y narran. Reconoce los patrones sin necesidad de vivirlos.

⚙️ ¿Qué lo hace “de gran escala”?

Cuando decimos “Large (gran escala)”, no es una metáfora: nos referimos al tamaño de su arquitectura.

Un LLM como GPT-4 puede tener billones de parámetros, que son los pequeños componentes ajustables que determinan cómo interpreta y genera lenguaje. Cuantos más parámetros tiene, más capacidad tiene para identificar matices, resolver ambigüedades, mantener el estilo, y adaptarse al contexto.

Esto es comparable a una orquesta:

Cuantos más instrumentos bien afinados tienes, más rica y compleja puede ser la música que interpretas.

 ¿Qué ocurre cuando haces una pregunta?

Cuando escribes algo como:

"¿Cuál es la diferencia entre inteligencia emocional e inteligencia lógica?"

Esto es lo que sucede en segundos:

1. Tu texto se convierte en tokens.
2. El modelo analiza cada token dentro del contexto general.
3. Predice, uno por uno, cuál debería ser el siguiente token para construir una respuesta coherente, clara y útil.
4. Arma la frase final, y te la entrega en lenguaje humano.

Y todo esto sucede en menos tiempo del que te toma respirar profundo.

 ¿Por qué esto es tan revolucionario?

Porque un LLM no está programado para tareas específicas. No fue diseñado solo para traducir, o solo para responder preguntas. Fue entrenado para comprender el lenguaje como un todo, lo cual le permite adaptarse a infinitos usos:

- Escribir guiones, correos o discursos
- Resolver dudas académicas
- Corregir textos o reescribirlos en otro tono
- Traducir entre idiomas
- Ayudarte a programar, escribir código y detectar errores
- Simular personajes, escribir novelas o colaborar en videojuegos
- Actuar como tutor, mentor o asistente virtual

Es una especie de copiloto digital del conocimiento. Y cuanto mejor lo entendamos, más valor podremos sacarle.

 En resumen

Un LLM es mucho más que un modelo matemático:

Es el resultado de años de investigación, toneladas de datos, y el deseo de construir una tecnología capaz de comunicarse con nosotros en nuestro idioma, sin traducirlo a código, fórmulas o comandos.

No es una mente humana, pero ha aprendido a hablar como una.

No tiene ideas propias, pero puede ayudarnos a desarrollar las nuestras.

No siente, pero puede escribir sobre sentimientos de forma impactante.

Y lo más importante: está aquí para ampliar lo que somos capaces de hacer, no para reemplazarlo.

5.2.2 Tokens: la unidad del lenguaje computacional

Si los humanos pensamos en ideas, oraciones y emociones, las inteligencias artificiales lo hacen a través de fragmentos más pequeños, casi microscópicos,

llamados tokens. Esta es una de las claves menos visibles —pero más importantes— para entender cómo funciona un LLM.

Podríamos decir que los tokens son el lenguaje dentro del lenguaje. Son la materia prima con la que la inteligencia artificial construye frases, genera respuestas, escribe poesía o redacta un código.

🧩 ¿Qué es exactamente un token?

Un token es una unidad mínima de texto que un modelo de lenguaje utiliza para procesar información. Pero a diferencia de una palabra, no tiene un tamaño fijo.

Un token puede ser:

- Una palabra completa: "gato"
- Una parte de una palabra: "intelligen-"
- Un sufijo o prefijo: "-cia", "anti-"
- Un signo de puntuación: ".", "!"
- Un espacio o símbolo: " ", "@", "#"

En lugar de leer frases como lo hacemos los humanos, el modelo divide cada oración en pequeñas piezas que puede analizar, comparar y predecir.

🧠 ¿Cómo funciona este sistema?

Veámoslo con un ejemplo práctico.

Tú escribes:

"La inteligencia artificial está transformando el mundo."

Para ti, es una idea clara, fluida y completa.

Pero para un LLM, esto se convierte en algo como:

`["La", "inteligencia", "artificial", "está", "transformando", "el", "mundo", "."]`

Cada uno de esos bloques es un token, y es lo que el modelo "lee".

Este proceso se llama tokenización: una traducción automática del lenguaje humano a un formato que la máquina pueda entender, procesar y predecir.

Es como si el modelo descompusiera cada oración en fichas de Lego, y luego las analizara, reorganizara o completara según lo que ha aprendido.

🎯 ¿Por qué los tokens son tan importantes?

- ♦ 1. Determinan cuánta información puede manejar el modelo

Cada LLM tiene una ventana de contexto, que se mide en tokens (no en palabras ni en caracteres).

Por ejemplo:

- GPT-3.5 puede procesar hasta 4.096 tokens en una conversación.

- Claude 3 puede trabajar con 200.000 tokens.
- Gemini 1.5 alcanza el millón.

Eso significa que cuanto más largos o complejos sean tus textos, más tokens ocuparán. Y si se supera el límite, el modelo comenzará a “olvidar” lo que había al principio.

♦ 2. Afectan el costo computacional

En plataformas como OpenAI o Anthropic, el uso de LLMs se cobra por tokens procesados.

Generar una respuesta corta puede costar unos cientos de tokens. Una respuesta extensa o muy detallada puede llegar a miles. Por eso, entender cómo se distribuyen los tokens ayuda a optimizar los costos si estás desarrollando una aplicación con IA.

♦ 3. Condicionan la precisión y coherencia del contenido generado

Un modelo no adivina palabras completas. Adivina el siguiente token más probable, dado el contexto anterior.

Es un sistema de predicción probabilística, donde el objetivo es construir la mejor secuencia posible, token por token.

🔍 ¿Cómo predice un token?

Imagina que escribes:

"La Tierra gira alrededor del..."

El modelo no tiene una respuesta grabada. Lo que hace es analizar todos los contextos similares que ha “leído” durante su entrenamiento y asignar una probabilidad matemática a las posibles continuaciones:

- "Sol" → 92%
- "centro" → 3%
- "eje" → 2%
- "amor" → 0.1%

Con esa probabilidad, elige “Sol” como la continuación más coherente.

Y una vez que lo hace, repite el proceso para el siguiente token... y el siguiente... y así sucesivamente, hasta que construye una respuesta completa.

Así es como el modelo puede mantener el hilo de una historia, responder una pregunta o componer una canción.

📏 ¿Cuántos tokens hay en un texto?

Una pregunta frecuente es: *“¿Cómo sé cuántos tokens tiene lo que escribí?”*

Aunque depende del idioma y del modelo, hay una regla práctica:

- En inglés, 1.000 palabras = ~750 tokens
- En español, 1.000 palabras = ~1.100 tokens (por ser un idioma más largo en estructura)

Esto significa que un modelo con una ventana de 100.000 tokens podría procesar un libro entero de 80–90 páginas, dependiendo del contenido.

Algunas herramientas como [OpenAI Tokenizer](#) permiten pegar un texto y ver cuántos tokens ocupa.

 ¿Cómo afectan los tokens al desarrollo de apps con IA?

Si vas a construir una aplicación, un asistente virtual, una plataforma educativa o una función automatizada, entender cómo se usan los tokens te permite diseñar mejor la experiencia:

- Puedes dividir tareas largas en bloques tokenizados.
- Evitarás cortes abruptos en las respuestas.
- Controlarás mejor los tiempos de respuesta y los costos.
- Sabrás cómo adaptar la interacción según el modelo y su capacidad.

 En resumen

Los tokens son el lenguaje secreto con el que la inteligencia artificial entiende y construye nuestras palabras.

Son diminutas piezas de texto que, una vez unidas, pueden crear respuestas brillantes, poemas emocionantes, soluciones complejas o conversaciones naturales.

No los vemos, pero están detrás de todo lo que ocurre cuando interactuamos con un LLM.

Y cuanto más entendamos su lógica, más preparados estaremos para usar esta tecnología de forma consciente, estratégica y creativa.

Porque al final, la calidad de lo que obtenemos no depende solo de lo que preguntamos... sino también de cómo se construyen las respuestas, token por token.

5.2.3 Ventana de Contexto (Context Window)

Imagina que estás teniendo una conversación profunda con alguien. Mientras más recuerde esa persona sobre lo que ya hablaste —tu nombre, tus ideas previas, tus dudas anteriores—, más coherente y significativa será su respuesta. En el mundo de la inteligencia artificial, esa “memoria activa” se llama ventana de contexto.

La ventana de contexto (o *context window*) es, simplemente, la cantidad de información que un modelo puede considerar al mismo tiempo cuando genera una respuesta. No se mide en párrafos ni en palabras, sino en tokens, que como vimos antes, son las unidades mínimas de lenguaje para una IA.

 ¿Por qué es tan importante esta ventana?

Porque marca la diferencia entre una conversación superficial y una conversación profunda.
Entre una IA que se pierde en el camino, y una que “recuerda” lo que estás tratando de hacer.

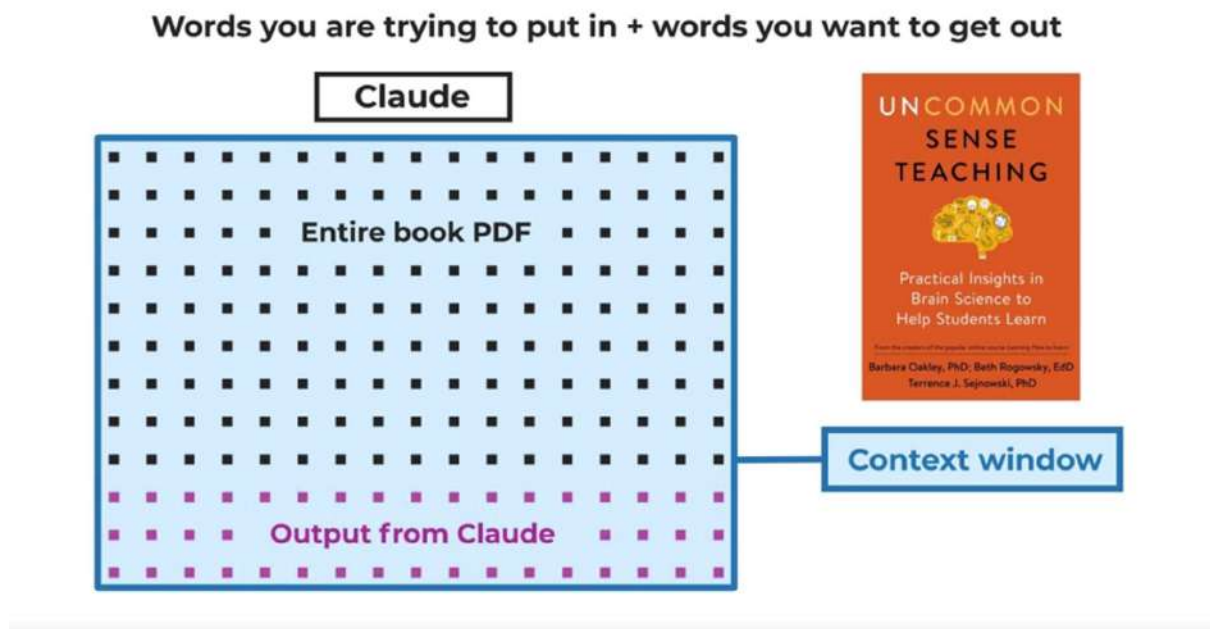


Figura 5.13 Ventana de Contexto

Una ventana de contexto pequeña es como hablar con alguien que se olvida de lo que dijiste hace dos minutos.

Una ventana de contexto amplia es como trabajar con un asistente que puede revisar todo un documento, analizarlo, resumirlo, y aún así recordar tus instrucciones iniciales.

🧠 ¿Cuánto puede recordar cada modelo?

Aquí es donde las cosas se ponen interesantes. No todos los modelos de IA tienen la misma capacidad para mantener información activa. 📖 ¿Qué permite una ventana de contexto amplia?

Cuando hablamos de 100.000 o 1.000.000 de tokens, estamos hablando de una IA capaz de:

- Leer y analizar libros enteros en una sola pasada.
- Comparar múltiples documentos legales, técnicos o científicos sin perder el hilo.
- Recordar detalles de una conversación larga (como una asesoría educativa de varias horas).
- Comprender mejor el tono, el estilo y la intención de quien escribe.
- Mantener consistencia narrativa en guiones largos, ensayos o novelas.

- Traducir, resumir o responder preguntas sobre documentos complejos sin necesidad de fragmentarlos.

Veamos algunos ejemplos reales de distintos modelos actuales

Tabla 5.4. Máxima Ventana de Contexto por Modelo

Modelo	Máxima ventana de contexto	¿Qué significa esto en la práctica?
GPT-3.5 (OpenAI)	4.096 tokens	Ideal para conversaciones cortas o tareas simples. Suele perder el hilo en documentos largos.
GPT-4 (OpenAI)	8.192 tokens (standard)	Mejor capacidad de retención, pero aún con límites en textos extensos.
GPT-4 Turbo (OpenAI)	128.000 tokens	Puede trabajar con libros completos, análisis extensos, conversaciones largas y documentación técnica compleja.
Claude 2 (Anthropic)	100.000 tokens	Excelente para procesamiento de múltiples documentos o materiales largos. Muy usado en educación y empresas.
Claude 3 Opus	200.000 tokens	Uno de los modelos con mayor capacidad de contexto del mercado actual. Puede trabajar con libros enteros, bases de datos textuales y más.
Gemini 1.5 Pro (Google)	1.000.000 tokens	Capacidad masiva de lectura y retención. Aún en pruebas, pero representa el futuro de la comprensión profunda.
Mistral 7B	8.192 tokens	Modelo ligero, útil para tareas básicas o integración en apps personalizadas.
Mixtral (Mistral)	32.000 tokens	Mezcla de expertos, buena para análisis de textos más largos. Aún optimizado para tareas específicas.
LLaMA 2 (Meta)	4.096 – 32.000 tokens	Varía según la versión y la implementación. Muy usado en entornos académicos y proyectos de código abierto.

⚠ ¿Tiene límites esta memoria?

Sí, y es importante entenderlo.
La ventana de contexto no es una memoria a largo plazo, sino una memoria temporal. El modelo recuerda solo lo que entra dentro de ese espacio. Si te pasas del límite, empieza a “olvidar” lo primero que dijiste para poder acomodar lo nuevo.

Además, cuanto más grande es la ventana de contexto, más costosa es la operación (en términos de procesamiento y precio, si estás usando una API de pago).

🎯 ¿Cómo usar estratégicamente la ventana de contexto?

Aquí algunos consejos prácticos:

1. Divide en bloques lógicos: Si usas un modelo con ventana corta, fragmenta el texto o conversación en partes más pequeñas y trabaja una a la vez.
2. Usa títulos y resúmenes: Introducir un resumen o subtítulos ayuda a que el modelo “comprime” mejor la información.
3. Elige el modelo correcto: No uses GPT-3.5 si necesitas procesar 100 páginas. Opta por GPT-4 Turbo, Claude 3 o Gemini.
4. Apóyate en herramientas híbridas: Algunas plataformas combinan IA con bases de datos que permiten búsquedas semánticas fuera de la ventana de contexto.

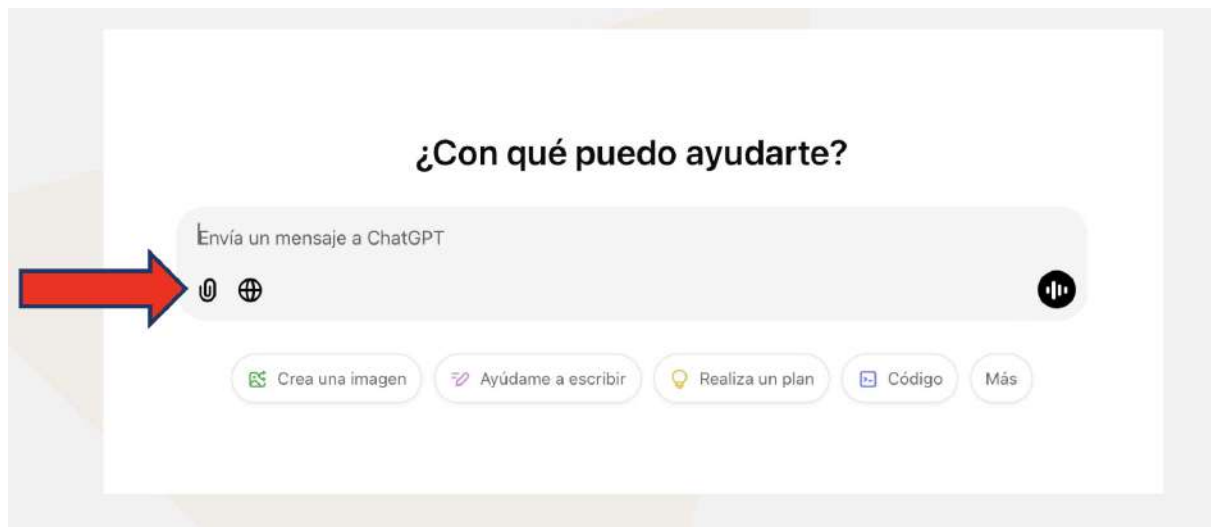


Figura 5.14. Como subir un documento en ChatGPT

🧩 El futuro: ventanas infinitas

Empresas como Google DeepMind, Anthropic y OpenAI están trabajando en modelos capaces de procesar no solo millones de tokens, sino también de conectar la IA a memorias externas, bases de conocimiento vivas o interacciones anteriores.

Estamos avanzando hacia modelos que no solo “recuerdan” más, sino que aprenden contigo.

Modelos que pueden seguir un proyecto durante semanas, colaborar contigo en investigaciones extensas o ayudarte a escribir un libro sin perder el hilo.

✨ En resumen

La ventana de contexto es mucho más que una característica técnica. Es lo que define cuán profunda, coherente y útil puede ser una IA en una conversación o tarea.

- Una ventana pequeña sirve para respuestas rápidas.
- Una ventana amplia permite construir ideas complejas.
- Una futura “memoria expandida” hará que la IA no solo responda... sino que te acompañe en el proceso creativo, paso a paso.

Y ese, sin duda, será un nuevo capítulo en la historia de la inteligencia artificial,

5.2.4 APIs: llevar la IA a donde tú quieras

La mayoría de las personas piensan en la inteligencia artificial como una conversación en una ventana de chat. Una caja mágica donde escribes algo, presionas “Enter” y obtienes una respuesta brillante. Y aunque eso es cierto, es apenas la punta del iceberg.

Detrás de ese intercambio aparentemente sencillo, existe una tecnología que hace posible que los modelos de lenguaje interactúen con otras aplicaciones, plataformas y sistemas. Esa tecnología se llama API.

✿ ¿Qué es una API, realmente?

Una API (Application Programming Interface, o Interfaz de Programación de Aplicaciones) es como una especie de puente invisible entre dos sistemas. Permite que una aplicación “hable” con otra para pedirle algo o entregarle información.

En el caso de los LLMs, la API es la vía por la cual una plataforma externa puede conectarse con la inteligencia artificial y pedirle que haga algo: escribir un texto, resumir un documento, responder una pregunta, traducir una idea, generar un correo, analizar una imagen... y mucho más.

⚙️ ¿Cómo funciona una API de IA generativa?

Pongamos un ejemplo simple:

1. Imagina que tienes una aplicación para aprender idiomas.
2. Quieres que los estudiantes puedan hacerle preguntas al estilo: “¿Cómo se dice ‘volveré pronto’ en francés?”
3. En lugar de escribir tú todas las respuestas posibles, conectas tu app con una API de IA.
4. Cada vez que el estudiante escribe su pregunta, la app la envía a través de la API hacia un LLM.
5. El modelo procesa el texto, genera una respuesta, y la devuelve en segundos.

Resultado: Tu app ahora tiene un asistente inteligente, multilingüe, siempre disponible.

¿Qué se puede hacer gracias a estas APIs?

La verdadera revolución está aquí: en la capacidad de llevar la IA a donde tú quieras, más allá del navegador, más allá del chat. Estos son algunos de los usos más comunes (y poderosos):

Educación y e-learning

- Crear tutores virtuales personalizados, que se adaptan al nivel y estilo de cada estudiante.
- Generar cuestionarios automáticos, resúmenes de clase, fichas de estudio o explicaciones paso a paso.
- Analizar el progreso del alumno y sugerir actividades según su desempeño.

Atención al cliente y automatización

- Desarrollar asistentes conversacionales 24/7, capaces de resolver dudas, gestionar pedidos o canalizar reclamos.
- Automatizar la redacción de respuestas personalizadas por correo o chat.
- Clasificar e interpretar tickets de soporte para ahorrar tiempo humano.

Empresas y productividad interna

- Integrar la IA en CRM, ERPs o plataformas de gestión para resumir reportes, redactar propuestas o asistir en tareas administrativas.
- Mejorar procesos de reclutamiento, escribiendo descripciones de puestos o analizando currículums.
- Ayudar en la toma de decisiones con análisis predictivos y generación de insights.

Apps y productos digitales

- Incorporar funciones de IA en aplicaciones móviles: desde generadores de contenido hasta asistentes por voz.
- Desarrollar interfaces de usuario inteligentes que se adapten según lo que el usuario busca o necesita.
- Potenciar videojuegos con personajes que responden de forma dinámica gracias al lenguaje natural.

¿Qué ofrecen los principales proveedores?

Actualmente, hay varias empresas líderes que ofrecen APIs robustas y escalables para conectar con modelos de lenguaje avanzados:

◆ OpenAI

- Modelos como GPT-4 y GPT-4 Turbo
- Amplia documentación, gran comunidad de desarrolladores
- Uso comercial en plataformas como Notion, Duolingo, Canva, Microsoft, entre otras

♦ Anthropic

- Modelos como Claude 2 y Claude 3 (con ventanas de contexto enormes)
- Foco en la alineación ética y respuestas seguras
- Muy usado en sectores educativos y empresariales por su capacidad para procesar documentos largos

♦ Google (Gemini API)

- Integración con sus servicios (Gmail, Docs, Drive)
- Capacidades multimodales (texto, imágenes, video)
- Avances experimentales con memorias a largo plazo

♦ Cohere y Mistral

- Modelos de código abierto y privado
- Ideales para quienes buscan personalizar sus implementaciones
- Gran flexibilidad para entrenamiento y ajustes a medida

¿Por qué las APIs son tan valiosas?

Porque permiten algo fundamental: escalar.

Un equipo humano no puede escribir miles de respuestas personalizadas por día, ni generar contenido para 100 clientes a la vez, ni procesar manualmente todas las solicitudes de usuarios.

Una IA conectada por API sí puede.

Además, liberan tiempo humano para tareas de mayor valor, permiten ofrecer experiencias personalizadas a gran escala, y reducen el margen de error humano en tareas repetitivas.

¿Y la seguridad? ¿Qué pasa con los datos?

Las APIs modernas están diseñadas con protocolos de seguridad avanzados. La mayoría permiten configurar niveles de acceso, anonimización de datos y almacenamiento temporal.

No obstante, es clave:

- Leer las políticas de privacidad de cada proveedor
- Evitar enviar datos sensibles sin encriptación
- Usar entornos seguros, especialmente en sectores como salud o educación

También es importante señalar que algunos proveedores permiten ejecutar modelos de lenguaje en servidores propios(on-premise), lo cual garantiza aún más control sobre los datos.

✨ En resumen

Las APIs son el puente que convierte a la IA generativa en una fuerza útil, integrada, ubicua.

Gracias a ellas, la IA deja de ser una herramienta aislada para convertirse en una parte viva de tus procesos, productos y experiencias digitales.

Y lo mejor de todo: no necesitas ser un experto en inteligencia artificial para usarlas. Solo hace falta una idea clara y la decisión de integrar esta nueva inteligencia en tu mundo.

Porque cuando conectas un modelo con una aplicación, ya no solo estás automatizando... estás potenciando el futuro.

5.2.5 ¿Por qué algunos modelos son cerrados y otros de código abierto?

En el universo de la inteligencia artificial, no todo es blanco o negro, pero hay una distinción que marca profundamente cómo se construye, se accede y se controla la tecnología: la diferencia entre modelos de código cerrado y modelos de código abierto.

Esta diferencia no es solo técnica. Tiene implicaciones éticas, económicas, estratégicas y hasta ideológicas. Porque detrás de cada elección tecnológica, hay una forma de entender el acceso al conocimiento, la privacidad, la soberanía digital y el futuro de la innovación.

🧱 ¿Qué significa que un modelo sea cerrado o abierto?

🔒 Modelo de código cerrado (closed-source)

Un modelo cerrado es como una caja negra. Puedes usarlo, pero no puedes ver qué hay dentro. No sabes con qué datos fue entrenado, ni tienes acceso al código que lo hace funcionar. Todo está gestionado por la empresa que lo creó.

Ejemplos populares:

- GPT-4 / GPT-4 Turbo (OpenAI)
- Claude (Anthropic)
- Gemini (antes Bard) (Google)

Estos modelos suelen estar disponibles a través de APIs o interfaces web, y funcionan como servicios en la nube: tú haces una consulta, y la IA responde, pero todo el procesamiento ocurre en los servidores del proveedor.

🔓 Modelo de código abierto (open-source)

Un modelo abierto es como una caja de herramientas. Puedes descargarlo, estudiarlo, modificarlo e incluso entrenarlo con tus propios datos. Está disponible para que cualquier persona —o institución— lo adapte a sus necesidades.

Ejemplos destacados:

- LLaMA 2 y 3 (Meta)
- Mistral 7B / Mixtral (Mistral)
- Falcon (Technology Innovation Institute, Emiratos Árabes)
- Dolly (Databricks)
- BLOOM (BigScience, comunidad internacional)

Estos modelos suelen estar disponibles en plataformas como Hugging Face, GitHub o repositorios especializados, y pueden ser ejecutados en servidores propios o en servicios de nube compatibles.

Tabla 5.5. Diferencias clave entre modelos cerrados y abiertos

Aspecto	Modelos Cerrados	Modelos Abiertos
Acceso al código	No	Sí
Transparencia del entrenamiento	Limitada o nula	Alta (según el modelo)
Control sobre los datos	Depende del proveedor	Total (si se usa en infraestructura propia)
Flexibilidad	Baja: se adapta a través de la API	Alta: se puede modificar y entrenar
Facilidad de uso	Muy alta (plug & play)	Requiere conocimientos técnicos
Privacidad y seguridad	Depende de políticas del proveedor	Se puede garantizar en servidores privados
Costo	Pago por uso (tokens/API)	Gratuito o costo único (infraestructura)
Actualización y mantenimiento	A cargo del proveedor	A cargo del usuario o comunidad

 **¿Cuál conviene más? Todo depende del contexto**

 **Cuándo elegir un modelo cerrado**

- Cuando necesitas una solución rápida y confiable.
- Si no tienes equipo técnico propio o tiempo para configurar un modelo desde cero.

- Si estás desarrollando una app sencilla, chatbot o herramienta de productividad.
- Cuando buscas máxima precisión y rendimiento desde el primer uso.

Por ejemplo: un profesor que quiere implementar IA en su aula a través de una plataforma tipo Notion o Canva puede usar GPT-4 o Claude vía API sin complicaciones.

✓ Cuándo elegir un modelo abierto

- Cuando necesitas mayor control sobre los datos (ej. salud, gobierno, investigación).
- Si deseas personalizar el comportamiento del modelo con datos propios.
- Cuando buscas independencia tecnológica y evitas depender de una sola empresa.
- Si tu institución tiene un equipo técnico y recursos de infraestructura (nube o servidores propios).

Por ejemplo: una universidad que quiere desarrollar un asistente académico entrenado con sus propios contenidos, exámenes y documentos internos, probablemente se beneficie más de un modelo abierto como LLaMA o Mixtral.

✚ Una cuestión de valores (y de estrategia)

Más allá de lo técnico, esta elección también implica una postura frente a la democratización de la IA.

- Los defensores del código abierto apuestan por la transparencia, la soberanía digital y el acceso equitativo al conocimiento.
- Las empresas de modelos cerrados priorizan la seguridad, el control de calidad, la optimización a gran escala y la protección ante usos indebidos.

Ambos enfoques tienen mérito. Ambos impulsan el avance tecnológico desde ángulos distintos. Y en muchos casos, lo ideal no es elegir uno en contra del otro, sino combinarlos inteligentemente.

🧠 Un futuro híbrido

Cada vez más organizaciones están optando por estrategias mixtas:

- Usan modelos cerrados para tareas de front-end (como generar contenido)
- Y modelos abiertos para analizar datos internos, mantener privacidad o entrenar con información sensible.

Además, la tendencia del “fine-tuning” (ajuste fino) está haciendo que los modelos abiertos se vuelvan cada vez más competitivos, ya que pueden adaptarse de forma precisa a sectores específicos: medicina, derecho, educación, industria, etc.

✨ En resumen

La diferencia entre código abierto y cerrado no es solo una elección técnica, es una decisión estratégica.

- Y si estás construyendo el futuro con IA, tal vez la mejor respuesta no esté en los extremos... sino en cómo combinas lo mejor de cada mundo.

El desarrollo de los modelos de lenguaje de gran escala (LLMs) se ha convertido en una nueva forma de geopolítica tecnológica. Ya no se trata solo de escribir textos o responder preguntas, sino de definir quién tiene el control sobre la infraestructura cognitiva del siglo XXI.



Estados Unidos, Europa y Asia libran una carrera silenciosa —pero estratégica— por liderar esta revolución. Los países no solo compiten por lanzar el modelo más potente, sino también por crear ecosistemas de inteligencia artificial adaptados a sus valores, necesidades e intereses comerciales o estatales.

Tabla 5. 6 Resumen Actualizado de Empresas, Modelos de Lenguaje, Usuarios y Fecha (2025)

Empresa	Modelo de Lenguaje	No. Usuarios (millones)	Fecha
Google	Gemini 2.5 Flash	350	Marzo 2025
Amazon (Anthropic)	Claude 3.5 Sonnet	25	Junio 2024
Microsoft (OpenAI)	ChatGPT 4.0 (GPT-4o)	800	Abril 2025
Meta (Facebook)	LLaMA 3.1	300	Julio 2024
Baidu	ERNIE 4.5 Turbo	300	Abril 2025
Alibaba	Qwen 2.5 Max	50	Enero 2025
Tencent	Hunyuan 2.5	20–35	Abril 2025
DeepSeek	DeepSeek V3	100	Abril 2025

En esta sección, recorreremos los modelos y plataformas más influyentes del mundo, con foco especial en:

- Su arquitectura tecnológica
- Sus ventajas diferenciales
- Su aplicación educativa y profesional
- Su papel en el contexto geopolítico y cultural

5.3.1. LLMs en Estados Unidos: el epicentro de la innovación

Estados Unidos lidera actualmente la carrera global en el desarrollo de LLMs. Este dominio responde a una convergencia poderosa: talento técnico, capital de riesgo, universidades de élite, infraestructura digital y un ecosistema empresarial ágil.

Los modelos nacidos en Silicon Valley, Seattle o San Francisco no solo marcan el ritmo de la innovación, sino que están redefiniendo cómo enseñamos, aprendemos, creamos y trabajamos.

1. OpenAI (ChatGPT)

- Modelo: GPT-3.5 / GPT-4 / GPT-4 Turbo

- Alianzas: Microsoft (Azure)
- Fortalezas:
 - Conversación natural, coherencia contextual
 - Capacidad multimodal (texto, imagen, código)
 - Integración profunda con Microsoft Office, Copilot y Azure OpenAI



Figura 5.16 <https://openai.com/es-419/chatgpt/overview/>

- Aplicaciones:
 - Educación a distancia, corrección automática de tareas, generación de contenidos académicos, asistentes de programación y escritura creativa

Tabla 5. 7. Evolución de los Modelos de ChatGPT

Versión	Lanzamiento	Tokens contextuales	Capacidades clave
GPT-2	2019	1,024	Generación básica de texto
GPT-3	2020	2,048	Mejora de fluidez, razonamiento básico
GPT-3.5	2022	4,096	Uso en ChatGPT gratuito
GPT-4	2023	8,192 – 32,000	Razonamiento complejo, generación precisa
GPT-4 Turbo	2023	Hasta 128,000	Análisis de libros, codificación avanzada

GPT-4 Turbo representa el punto más avanzado de la línea GPT, ideal para tareas de alto volumen, procesos educativos largos o generación multimodal (texto, código, imágenes).

OpenAI es el nombre más reconocido en IA generativa. ChatGPT no solo se ha masificado, sino que ha cambiado la percepción pública sobre lo que una IA puede hacer. Su flexibilidad lo convierte en una herramienta transversal.

Perfil global de los usuarios de ChatGPT

La inteligencia artificial generativa no solo está cambiando cómo se enseña, se escribe o se programa. También está transformando **quiénes la usan, cómo la usan y con qué propósito**. Comprender el perfil de los usuarios de ChatGPT es fundamental para anticipar el impacto social, cultural y educativo de esta tecnología en los próximos años.

Aquí presentamos un análisis narrativo sobre su expansión global y la evolución del comportamiento de sus usuarios.

1. Tiempo récord para alcanzar el primer millón de usuarios

Cuando OpenAI lanzó ChatGPT en noviembre de 2022, pocos imaginaban que estaban frente a **la aplicación de más rápido crecimiento en la historia digital**. Solo pasaron **cinco días** para que la herramienta alcanzara su primer millón de usuarios. Cinco. Días.



Figura 5.17 Tiempo para llegar a 1 millón de usuarios

Para ponerlo en perspectiva:

- Instagram tardó **2,5 meses**,
- Spotify, **5 meses**,
- Netflix, **3,5 años**.

Este fenómeno marcó **el inicio de una nueva era de adopción tecnológica acelerada**, donde los usuarios no esperaban tutoriales ni certificaciones: simplemente entraban, escribían... y descubrían el poder de una IA conversacional que entendía, respondía y generaba.

Este fenómeno marcó **el inicio de una nueva era de adopción tecnológica acelerada**, donde los usuarios no esperaban tutoriales ni certificaciones: simplemente entraban, escribían... y descubrían el poder de una IA conversacional que entendía, respondía y generaba.

El interés masivo no surgió por campañas de marketing agresivas, sino por **una combinación perfecta de viralidad, utilidad y asombro**. ChatGPT se volvió un fenómeno cultural, educativo y profesional, todo al mismo tiempo.

2. Cantidad actual de usuarios

A medida que ChatGPT evolucionó de la versión 3.5 a la 4.0 (y próximamente a la 5.0), su base de usuarios creció de forma explosiva. Para abril de 2025, se estima que la plataforma ya cuenta con **más de 800 millones de usuarios activos semanales en todo el mundo**.

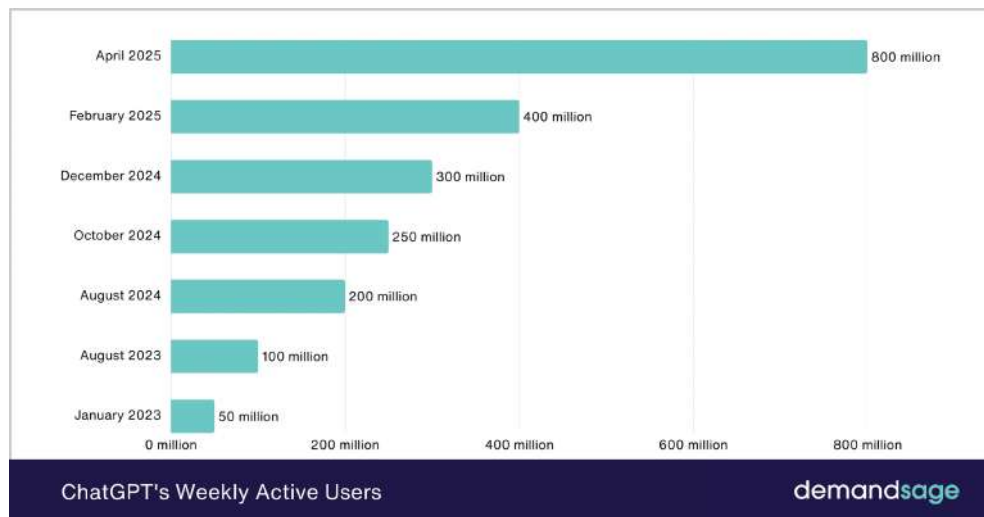


Figura 5.18 No. de usuarios de ChatGPT

Este número refleja mucho más que una moda pasajera. Significa que **cientos de millones de personas usan inteligencia artificial regularmente** para:

- redactar documentos académicos,
- traducir y corregir textos,
- generar ideas de negocio,

- resolver problemas técnicos,
- aprender nuevos idiomas o temas complejos.

Lo que comenzó como una curiosidad se ha transformado en **una herramienta esencial del día a día digital**. ChatGPT ya no es solo un "chat inteligente". Para muchos, es **una extensión de su pensamiento, su productividad y su forma de aprender**.

3. Distribución por edad

Uno de los aspectos más fascinantes del fenómeno ChatGPT es su penetración generacional. Aunque su diseño es universal, **la mayoría de sus usuarios pertenecen a generaciones jóvenes**, nativos digitales que han crecido interactuando con tecnología.

Distribución por edades:

- **18 a 24 años:** 28,04%
- **25 a 34 años:** 33,52%
- **35 a 44 años:** 18,35%

Esto significa que **más del 60% de los usuarios tienen menos de 35 años**, una señal clara de que la IA está siendo adoptada por quienes definen las tendencias, el lenguaje digital y los entornos de trabajo del futuro.

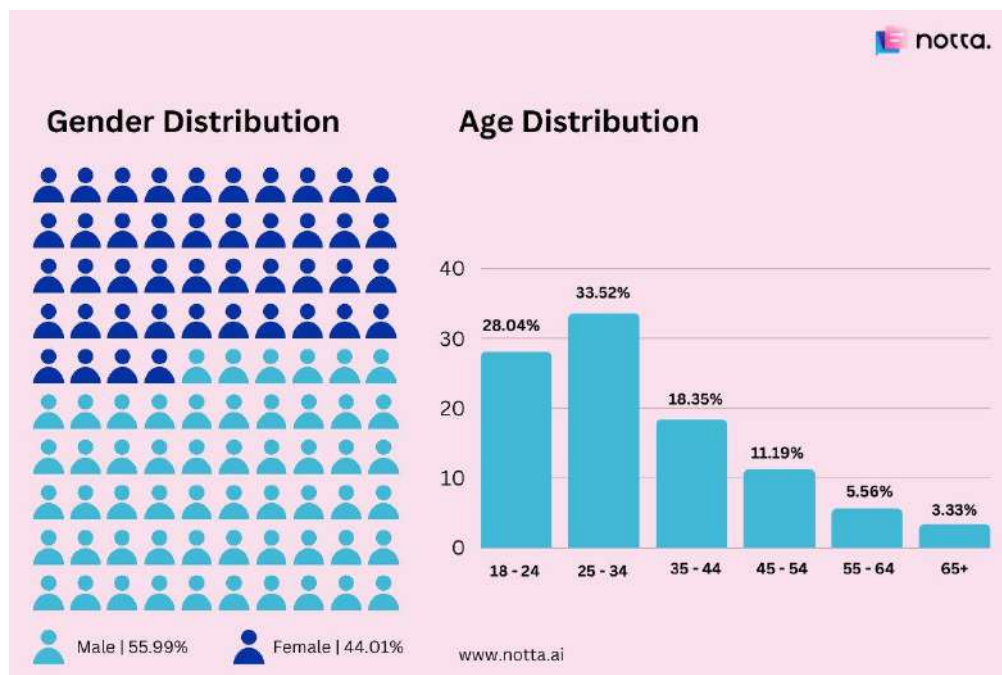


Figura 5.19 Distribución de usuarios de ChatGPT por edad

Pero también plantea un desafío:

👉 ¿Qué pasa con las generaciones mayores?

👉 ¿Cómo garantizamos la inclusión digital de quienes no han crecido con este tipo de herramientas?

El reto de la alfabetización en IA ya no es técnico, es generacional. Y es urgente.

 4. Distribución geográfica

ChatGPT no es una tecnología de nicho ni regional. Es **una herramienta global**, con una base de usuarios distribuida en casi todos los continentes. Aun así, existen países donde la adopción ha sido significativamente más alta, configurando **núcleos de uso intensivo de IA**.

Principales países por porcentaje de usuarios:

Según los datos más recientes de Similarweb correspondientes a febrero de 2025, la distribución del tráfico global hacia ChatGPT por país es la siguiente:

Tabla 5.8 Distribución de Usuarios de ChatGPT por país

País	Porcentaje de visitas
Estados Unidos	19.01%
India	7.86%
Brasil	5.05%
Canadá	3.57%
Reino Unido	3.48%
Otros países	61.03%

- Estos porcentajes reflejan la proporción de visitas al sitio web de ChatGPT desde cada país durante el mes de febrero de 2025.
- Es importante señalar que estas cifras pueden variar con el tiempo debido a factores como la adopción tecnológica, el acceso a internet y las políticas locales relacionadas con la inteligencia artificial.
- Para obtener información más detallada y actualizada sobre la distribución del tráfico por país, se puede consultar el análisis completo en Similarweb.

En América Latina, destacan especialmente **Brasil, Colombia y México**, con decenas de millones de visitas mensuales. Esto sugiere que la región está emergiendo como **un espacio fértil para la expansión de la IA generativa**, sobre todo en educación, emprendimiento digital y atención al cliente.

La expansión de ChatGPT no se mide solo en datos. Se mide en **transformaciones sociales, nuevas habilidades y desafíos éticos**. Comprender quién lo usa, cómo y dónde, **es el primer paso para diseñar políticas, programas educativos y estrategias digitales que respondan a la realidad del siglo XXI**.

La pirámide del dominio: la evolución del usuario de ChatGPT

La inteligencia artificial no solo evoluciona en términos de parámetros, velocidad o capacidad multimodal. **También transforma a quienes la utilizan.** La adopción de herramientas como ChatGPT ha generado un proceso de madurez progresiva entre los usuarios, desde la simple curiosidad hasta un dominio consciente y estratégico.

Esta pirámide ilustra cómo, al ritmo en que las versiones de ChatGPT avanzan, **también crecen las habilidades, la autoconciencia y el criterio de quienes la usan.** Cada peldaño representa no solo una etapa tecnológica, sino un cambio en la mentalidad del usuario frente a la IA generativa.



Figura 5.20 Evolución de la Competencia del usuario de la IA Generativa

🧩 Base de la pirámide

(Noviembre 2022 – ChatGPT 3.5)

Estado: Inconscientemente incompetente

“No sé lo que no sé sobre la IA.”

En esta etapa inicial, la mayoría de los usuarios **desconocía por completo el potencial y las limitaciones** de la inteligencia artificial generativa. Usaban ChatGPT como una curiosidad, una novedad viral, sin una comprensión clara de su funcionamiento ni de su alcance. La exploración era intuitiva, y el aprendizaje, casi accidental.

Segunda capa

(Junio 2023 – ChatGPT 3.5 Turbo / GPT-4)

Estado: Conscientemente incompetente

“Sé que existe, pero todavía no sé usarla bien.”

Con la llegada de GPT-4 y la integración de funciones más avanzadas, **los usuarios comenzaron a reconocer tanto el valor de la IA como sus propias limitaciones al usarla**. Aparecen las primeras guías, los tutoriales, los prompts especializados. Comienza la transición del uso superficial al uso estratégico, aunque todavía con dudas, errores y zonas grises.

Tercera capa

(Mayo 2024 – ChatGPT 4.0)

Estado: Inconscientemente competente

“Ya lo hago bien sin pensarlo mucho.”

Aquí, muchos usuarios han integrado la IA en su flujo de trabajo diario. **Ya no preguntan qué puede hacer la IA: la usan como una herramienta natural, casi automática**. La competencia técnica crece, las tareas se automatizan, y el lenguaje promptológico se vuelve parte del vocabulario cotidiano. El aprendizaje se vuelve implícito.

Cima de la pirámide

(¿2026? – ChatGPT 5.0 y más allá)

Estado: Conscientemente competente

“Sé exactamente lo que hago y cómo aprovecharla al máximo.”

Este nivel representa el objetivo ideal: **usuarios que dominan la IA con plena conciencia, criterio ético y estrategia aplicada**. Saben cómo entrenarla, cómo modular su comportamiento, cómo combinarla con otras herramientas. Aquí ya no hay improvisación: hay intención.

Es una cima aún marcada con un signo de interrogación, porque **muchos aún están en camino**. Pero con la llegada de modelos como GPT-5, más potentes, versátiles y personalizados, ese punto se vuelve cada vez más alcanzable.

Conclusión

La evolución no es solo tecnológica, sino **cognitiva y cultural**. No basta con que los modelos de IA mejoren. También necesitamos **usuarios preparados, conscientes y críticos**.

Porque, en última instancia, **la inteligencia más poderosa no será la artificial... sino la humana que sepa cómo usarla con propósito**.

ChatGPT-5: el próximo salto evolutivo de la inteligencia artificial conversacional

El desarrollo de modelos de lenguaje no se detiene. Tras la consolidación de **ChatGPT-4** como una herramienta versátil para redacción, análisis, codificación y asistencia educativa, **OpenAI ya perfila el lanzamiento de ChatGPT-5**, un modelo que promete transformar por completo nuestra relación con la inteligencia artificial.

No se trata de una simple actualización, sino de **una evolución estructural** que expande capacidades, escala técnica y visión funcional.

Un salto colosal en escala

Uno de los datos más impactantes filtrados hasta ahora es el aumento en la cantidad de parámetros:

- **ChatGPT-4:** ~75.000 millones de parámetros
- **ChatGPT-5:** se estima en **10.000 billones (10 trillones en escala anglosajona)**

 **Esto representa un crecimiento de 133 veces**, lo que permitirá procesar contextos más amplios, respuestas más precisas y una capacidad de razonamiento mucho más cercana al lenguaje humano.

Capacidades multimodales expandidas

GPT-5 dará un paso más allá del texto. Incluirá de forma nativa:

- Procesamiento de imágenes complejas (por ejemplo, interpretar diagramas o mapas conceptuales)
- **Análisis de video** (extraer datos clave de grabaciones educativas o demostrativas)
- **Posible integración de audio** para generar o comprender contenido hablado

 Esto convertirá al modelo en una herramienta transversal, capaz de **asistir desde el aula hasta el quirófano**, pasando por entornos creativos, legales o industriales

🧠 Memoria ampliada y comprensión contextual

Uno de los grandes límites actuales es la "memoria corta" de los modelos: olvidan partes anteriores de la conversación o pierden coherencia en interacciones largas.

GPT-5 incorporará un sistema de **memoria persistente y contextual**, permitiendo:

- Recordar información relevante de conversaciones anteriores
- Adaptar el tono y estilo según la relación con el usuario
- Generar respuestas más personalizadas y con seguimiento lógico

🔄 Este avance podría cambiar radicalmente la forma en que usamos la IA, haciéndola sentir menos como una máquina... y más como un **asistente personal con memoria y continuidad**.

🌍 Soporte multilingüe mejorado

Se espera que GPT-5 incluya mejoras sustanciales en:

- Traducción automática
- Comprensión de matices culturales y dialectos regionales
- Generación de contenido en más de 100 idiomas con mayor precisión

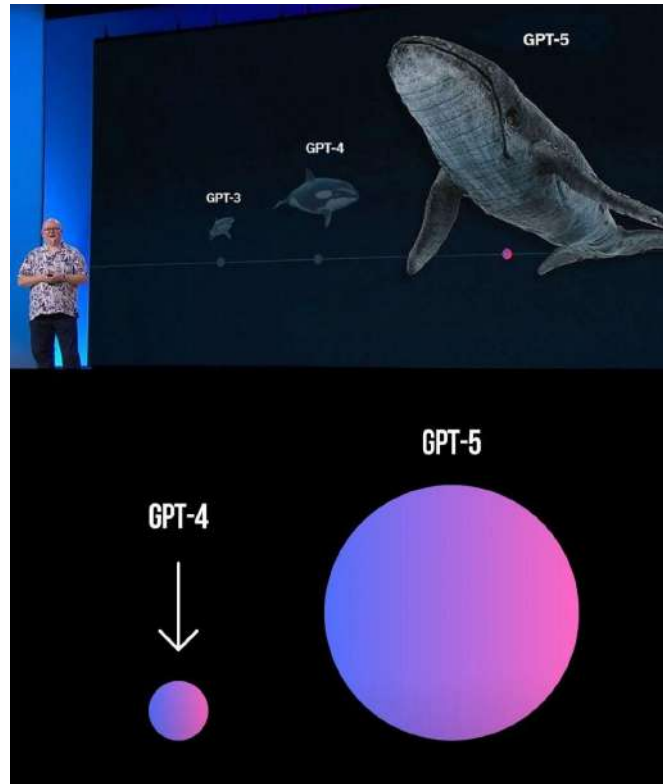


Figura 5.21 Comparación de GPT-4 vs GPT-5

Esto posiciona al modelo como una **plataforma ideal para educación multilingüe, medios internacionales y servicios públicos globales.**

Agentes autónomos de IA: hacia el trabajo sin intervención

Una de las apuestas más audaces de GPT-5 será el despliegue de **agentes autónomos**, capaces de operar con poca o ninguna supervisión humana.

Estos agentes podrían:

- Programar eventos, gestionar agendas y correos electrónicos
- Realizar compras o reservas online según preferencias aprendidas
- Automatizar tareas laborales complejas en múltiples entornos

 La IA dejará de ser solo una herramienta de consulta para convertirse en **un colaborador activo y autónomo** en nuestras rutinas.

Código y programación de nueva generación

Para desarrolladores y entornos técnicos, GPT-5 será aún más poderoso:

- Mejor comprensión semántica de múltiples lenguajes de programación
- Generación de funciones, depuración de errores, refactorización automatizada
- Soporte extendido a entornos de desarrollo y frameworks modernos

 GPT-5 no solo escribe código, **lo entiende, lo explica y lo mejora.**

Un modelo hacia la inteligencia general artificial (AGI)

Aunque GPT-5 aún no representa una AGI (Inteligencia General Artificial), muchos expertos creen que será el **modelo más cercano a una mente artificial versátil, autónoma y creativa** que hemos visto hasta ahora.

- Mayor capacidad de razonamiento
- Comprensión emocional más refinada
- Integración con sistemas externos para actuar en el mundo real

¿Cuándo y a qué costo?

Hasta ahora, no hay una fecha oficial de lanzamiento ni una confirmación de precios. Se espera que siga un modelo similar al actual:

- Versión gratuita con limitaciones (GPT-4.5 o GPT-3.5)
- Versión premium con acceso completo (suscripción mensual)

Tabla 5.9 Característica de GPT-4 versus GPT-5

Característica	GPT-4	GPT-5 (estimado)
Parámetros	~75 mil millones	10 billones (10,000,000,000,000)
Modalidad	Texto + imagen	Texto, imagen, video, audio
Memoria	Temporal (por sesión)	Persistente y contextual
Multilingüismo	Ampliado	Optimizado y culturalmente sensible
Agentes autónomos	No	Sí (en exploración inicial)
Código y programación	Muy bueno	Avanzado y estructurado

2. Google DeepMind (Gemini)

- Modelo: Gemini 1.5 (sucesor de Bard)
- Fortalezas:
 - Integración nativa con Google Search, Gmail, Docs, Drive
 - Capacidad multimodal avanzada
 - Actualización constante con acceso a la web
- Aplicaciones:
 - Trabajo colaborativo, análisis de datos, organización educativa en la nube, búsqueda avanzada

Google apuesta por convertir a Gemini en el asistente inteligente universal, aprovechando su ecosistema de productos. Su enfoque está en combinar IA con productividad en tiempo real.

3. Anthropic (Claude)

- Modelo: Claude 3.7 Sonnet
- Apoyo estratégico: Amazon
- Fortalezas:
 - Alta seguridad y ética en respuestas
 - Ventanas de contexto amplísimas (Claude 3 llega a 200.000 tokens)
 - Ideal para entornos regulados o con públicos sensibles
- Aplicaciones:

- Educación formal, asistencia empresarial, moderación de contenido, instituciones con enfoque de responsabilidad social

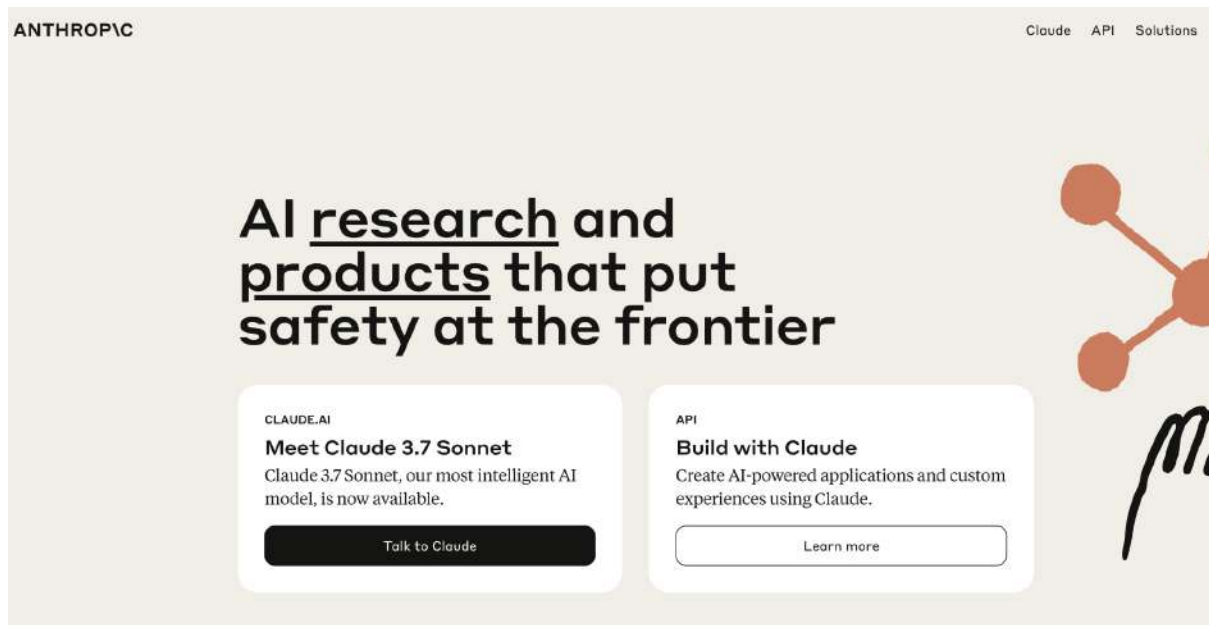


Figura 5.22. Claude 3.7 Sonnet <https://www.anthropic.com/>

Claude es una IA diseñada para ser confiable, incluso en conversaciones complejas o emocionalmente delicadas. Perfecta para docentes, consejeros, gestores de comunidad.

🚀 4. Meta (LLaMA)

- Modelo: LLaMA 2 / LLaMA 3 (open-source)
- Fortalezas:
 - Código abierto, ideal para personalización e investigación
 - Modelos ligeros, eficientes, adaptables
 - Amplio soporte en la comunidad investigadora
- Aplicaciones:
 - Creación de asistentes institucionales privados, investigación académica, desarrollo de plataformas educativas propias

Meta ha decidido democratizar la IA, facilitando su descarga y modificación. Su modelo es perfecto para universidades, ONGs y empresas que deseen adaptar su propio sistema con bajo costo.

🧩 5. xAI (Grok)

- Fundador: Elon Musk
- Modelo: Grok (integrado con la red social X)

- Fortalezas:
 - Respuesta en tiempo real, tono informal y provocador
 - Integración social (X/Twitter)
 - Recibe datos actualizados de eventos sociales en vivo
- Aplicaciones:
 - Interacción con usuarios, análisis de tendencias sociales, generación de contenido viral

Grok representa un enfoque disruptivo: una IA “irreverente” que responde con estilo sarcástico o informal, apuntando al entretenimiento y la cultura de redes.

6. LinkedIn (Pi – Personal Intelligence)

- Modelo: Pi
- Fortalezas:
 - Asistente profesional para el mercado laboral
 - Generación de contenido para CVs, publicaciones, cartas de motivación
 - Orientación laboral personalizada
- Aplicaciones:
 - Asesoría vocacional, preparación para entrevistas, mejora del perfil profesional

Pi busca posicionarse como el mentor de IA para el mundo del trabajo. Es un modelo enfocado en la empleabilidad y el desarrollo de carrera.

 Tabla 5. 10 Comparación de Modelos de IA Generativa

Modelo	Desarrollador	Punto Fuerte	Ideal para...
ChatGPT	OpenAI	Versatilidad y fluidez	Educación, redacción, programación
Gemini	Google	Integración con Google	Nube educativa, análisis de datos
Claude	Anthropic	Ética y seguridad	Docencia, salud, entornos controlados
LLaMA	Meta	Open-source y eficiencia	Investigación, desarrollo institucional
Grok	xAI	Tono informal y en tiempo real	Medios, redes sociales
Pi	LinkedIn	Enfoque profesional	Reorientación laboral, networking

El dominio estadounidense en el desarrollo de LLMs genera una ventaja competitiva considerable no solo para la industria tecnológica, sino también para las instituciones educativas que deseen incorporar estas herramientas a sus procesos de enseñanza-aprendizaje. Comprender las diferencias entre estas plataformas permite tomar decisiones informadas sobre su adopción, sus riesgos y sus beneficios en contextos pedagógicos.

5.3.2. LLMs en Europa: soberanía digital y propósito social

Mientras Estados Unidos marca el paso en innovación acelerada, Europa avanza con una brújula distinta: ética, soberanía tecnológica y enfoque social. El viejo continente no compite solo por velocidad, sino por sentido, regulación y autonomía. En lugar de una carrera desenfrenada, Europa apuesta por una IA al servicio de sus valores: derechos humanos, transparencia y acceso equitativo.

Los modelos europeos de LLM nacen en un ecosistema donde la innovación convive con marcos regulatorios estrictos, una ciudadanía crítica y un compromiso profundo con el interés público.

1. Aleph Alpha (Alemania)

Modelo: Luminous

Fortalezas:

- Multilingüismo avanzado (incluyendo idiomas minoritarios)
- Transparencia en el proceso de entrenamiento
- Cumplimiento con GDPR desde el diseño

Aplicaciones:

Gobiernos, universidades, instituciones europeas, empresas que manejan datos sensibles

Aleph Alpha busca crear una IA europea, ética y soberana, con infraestructura propia. Su enfoque en la explicabilidad y la interoperabilidad lo hace ideal para sectores donde la confianza es clave, como la salud o la administración pública.

2. Mistral (Francia)

Modelos: Mistral 7B / Mixtral 8x7B

Fortalezas:

- Código abierto y entrenamiento eficiente

- Modelos modulares (mezcla de expertos – MoE)
- Rendimiento competitivo con menos recursos

Aplicaciones:

Investigación, educación superior, desarrollo de herramientas propias en contextos académicos o empresariales

Mistral representa la versión europea del enfoque open-source, con una potente comunidad científica detrás. Su meta: democratizar el acceso a LLMs de alto rendimiento, sin depender de gigantes estadounidenses.

3. Noosa (Unión Europea – consorcio)

Modelo: Noosa LLM (en desarrollo)

Fortalezas:

- Enfoque colaborativo y paneuropeo
- Encriptación avanzada y privacidad diferencial
- Optimización para el sector público

Aplicaciones:

Servicios gubernamentales, procesos educativos nacionales, entidades que priorizan la protección de datos

Este modelo es fruto de la colaboración entre estados miembro. Más que competir con Silicon Valley, Noosa busca fortalecer la resiliencia digital europea, creando modelos que puedan ser auditados, adaptados y controlados localmente.

4. BLOOM (BigScience – iniciativa internacional liderada por Francia)

Modelo: BLOOM / BLOOMZ

Fortalezas:

- Transparencia total (datos, entrenamiento, parámetros)
- Multilingüismo (46 idiomas, incluyendo lenguas africanas y sudamericanas)
- Accesible y documentado para toda la comunidad científica

Aplicaciones:

Educación global, investigación, comunidades académicas en países en desarrollo

BLOOM es un símbolo de ciencia abierta, construida por cientos de investigadores de todo el mundo. Es una alternativa ética y diversa que pone la IA al alcance de las universidades, ONGs y colectivos sociales.

 Comparación rápida: ¿en qué se destacan?

Europa, a diferencia de EE.UU., no busca dominar el mercado, sino garantizar que la IA respete sus principios fundamentales. Sus modelos son herramientas estratégicas para preservar la autonomía tecnológica, proteger los derechos digitales y fomentar una IA centrada en el bien común.

Tabla 5.11 Comparación de Modelos IA en Europa

Modelo	Desarrollador	Punto fuerte	Ideal para...
Luminous	Aleph Alpha (Alemania)	Privacidad y multilingüismo	Gobiernos, salud, educación pública
Mistral	Mistral (Francia)	Eficiencia y código abierto	Investigación, universidades, startups
Noosa LLM	Consortio europeo	Seguridad y soberanía digital	Sector público, educación nacional
BLOOM	BigScience (internacional)	Ciencia abierta y diversidad cultural	Educación inclusiva, cooperación internacional

En un mundo cada vez más interconectado, conocer y entender los modelos europeos no solo es una cuestión técnica, sino también cultural y política. Adoptarlos puede significar apostar por una educación más justa, una ciencia más abierta y una innovación con propósito.

5.3.3. Retos y oportunidades de los LLM europeos

El camino de Europa en el desarrollo de LLMs no está exento de obstáculos, pero también está lleno de oportunidades únicas. Apostar por una IA responsable, ética y soberana implica tomar decisiones difíciles, pero con recompensas a largo plazo.

Retos principales

Fragmentación política y regulatoria

Europa no es un bloque homogéneo. Cada país tiene sus prioridades, idiomas, niveles de inversión y marcos legales. Esta diversidad, aunque enriquecedora, puede ralentizar el avance tecnológico conjunto.

Menor inversión privada en comparación con EE.UU.

El capital de riesgo europeo es más conservador. Esto limita el crecimiento

acelerado de startups de IA y dificulta la competencia directa con gigantes como OpenAI o Google.



Fuga de talento

Muchos científicos e ingenieros europeos migran hacia Silicon Valley, atraídos por salarios, recursos y visibilidad internacional. Esto debilita el ecosistema local de innovación.



Desconexión entre ciencia y mercado

A pesar de contar con excelentes centros de investigación, Europa a veces tarda en traducir sus descubrimientos en productos útiles, accesibles y escalables.



Oportunidades estratégicas



Marco ético y regulatorio sólido (IA Act)

Europa lidera el mundo en la creación de leyes para el uso responsable de la inteligencia artificial. El AI Act establece límites claros, fomenta la transparencia y protege los derechos ciudadanos. Esto podría convertirse en un estándar global.



IA multilingüe y multicultural

Frente a modelos anglocéntricos, los LLMs europeos están diseñados para reflejar la diversidad lingüística y cultural del continente. Esto amplía la inclusión educativa y social.



Apoyo institucional al sector público y la educación

Muchos modelos europeos están pensados desde el inicio para mejorar la administración pública, modernizar la educación y reducir la brecha digital. Esto abre un mercado institucional enorme y estable.



Colaboración abierta y ciencia compartida

Iniciativas como BLOOM demuestran que Europa puede liderar no desde la competencia cerrada, sino desde la co-creación internacional, inclusiva y transparente.



En resumen

Europa no corre en la misma pista que Estados Unidos, pero eso no significa que esté detrás. Su fuerza está en su visión de largo plazo, en su capacidad para equilibrar innovación con valores, y en su apuesta por una IA que sirva a las personas, no solo a los mercados.

Los LLM europeos quizás no sean los más rápidos en llegar al mercado, pero sí podrían ser los más confiables, adaptables y sostenibles a largo plazo. En un mundo

donde los datos son poder, Europa busca algo más valioso: confianza, equidad y futuro compartido.

5.3.4. LLMs en China: velocidad, escala y visión estratégica

China avanza en inteligencia artificial con una visión clara: no depender tecnológicamente de Occidente y liderar el desarrollo de IA como herramienta de modernización nacional. Este impulso está respaldado por enormes inversiones estatales, políticas públicas alineadas y un mercado interno de más de mil millones de usuarios.

La IA no es solo innovación: es parte de la narrativa del desarrollo económico, educativo y geopolítico del país. Y eso se refleja en la proliferación de modelos de lenguaje masivo (LLMs) desarrollados por empresas, universidades y consorcios nacionales.

Principales modelos de IA y LLM en China

1. DeepSeek-V3 (2024)

Desarrollador: DeepSeek (start-up con apoyo estatal)

Tipo: Open-source, Mixture of Experts

Ventana de contexto: 64,000 tokens

Parámetros: 236B (64 activos por inferencia)

Fortalezas:

- Razonamiento lógico y matemático avanzado
- Multilingüismo real (no solo chino-inglés)
- Código y pesos públicos (acceso abierto)

Tiemblan ChatGPT y Gemini: un chatbot chino con inteligencia artificial sacude Silicon Valley

Se trata de Deepseek V3, que generó un gran revuelo en el sector tecnológico, así como también en la bolsa de Wall Street.



Por Canal26

Viernes 3 de Enero de 2025 - 13:18



Inteligencia Artificial. Foto: Usplash

Figura 5.23. Lanzamiento de Deepseek V3 sacude Silicon Valley

DeepSeek-V3 ha generado impacto global. Sus capacidades rivalizan con GPT-4 y Claude 3, y su disponibilidad pública lo convierte en una opción real para el mundo académico y gobiernos fuera del eje occidental.

2. ERNIE Bot 4.0

Desarrollador: Baidu

Tipo: Multimodal, con integración total en el ecosistema Baidu

Fortalezas:

- Comprensión contextual profunda en chino
- Aplicaciones en salud, educación, logística
- IA aplicada a datos locales y búsqueda en tiempo real

ERNIE es el modelo más “institucional” de China. No solo responde a usuarios, sino que alimenta sistemas de infraestructura pública e industrial, desde hospitales hasta plataformas de enseñanza.

3. SparkDesk

Desarrollador: iFlytek

Enfoque: Reconocimiento de voz + LLM educativo

Fortalezas:

- Capacidades orales sobresalientes
- Educación personalizada por voz
- Integración con dispositivos móviles

Spark ha sido adoptado por más de 1.000 escuelas rurales. China lo ve como parte de su política de reducción de desigualdad educativa.



4. Tongyi Qianwen

Desarrollador: Alibaba Cloud

Modelos: Tongyi 1.0 y 2.0

Fortalezas:

- Integración con DingTalk y Aliyun
- Enfoque en negocios, logística, comercio electrónico
- Generación de documentos, correos, resúmenes automáticos

Alibaba busca que Tongyi sea el asistente empresarial de China, con funciones prácticas para PYMEs, educación en línea y gestión administrativa.



5. PanGu-Σ (Sigma)

Desarrollador: Huawei

Modelo: Pangu Alpha / PanGu-Sigma

Fortalezas:

- Potencia para aplicaciones científicas e industriales
- IA para ingeniería, simulación y predicción de materiales
- Uso en educación superior tecnológica

Huawei orienta su modelo no tanto al consumidor general, sino a sectores estratégicos: industria, energía, ciencia aplicada y universidades STEM.



6. Baichuan 2

Desarrollador: Baichuan Intelligence

Modelos: Baichuan-13B, Baichuan2-53B

Fortalezas:

- Modelos open-source optimizados
- Entrenamiento en gran escala multilingüe

- Alto rendimiento en tareas de lenguaje natural

Baichuan representa el ala más *open-source* de la IA china, ideal para universidades, centros de investigación y empresas que buscan personalización sin depender de proveedores extranjeros.

 **Tabla 5. 12 Comparativa: Modelos líderes en China**

Modelo	Desarrollador	Punto fuerte	Ideal para...
DeepSeek-V3	DeepSeek	Razonamiento avanzado, open-source	Ciencia, investigación, educación superior
ERNIE 4.0	Baidu	Integración con datos locales	Gobierno, salud, sistemas nacionales
SparkDesk	iFlytek	IA educativa por voz	Docentes, alumnos en zonas rurales
Tongyi Qianwen	Alibaba Cloud	Asistente corporativo y productivo	PYMEs, educación en línea, logística
PanGu-Sigma	Huawei	IA para industria y ciencia aplicada	Ingeniería, universidades, simulaciones
Baichuan 2	Baichuan Intelligence	Código abierto, eficiencia lingüística	Investigación, desarrollo institucional

 ¿Qué distingue al ecosistema chino de LLMs?

- Planificación estatal a largo plazo: La IA no es solo una moda, es parte del plan quinquenal nacional.
- Escala poblacional única: Cualquier piloto educativo o tecnológico alcanza millones de usuarios en poco tiempo.
- Open-source estratégico: China ofrece modelos libres no solo por altruismo, sino para ganar influencia en regiones que no confían en las Big Tech de EE.UU.
- Centrados en el idioma y cultura local: Mientras muchos modelos anglosajones “traducen” ideas, los modelos chinos las entienden desde su origen.

 IA China: entre eficiencia y control

Los LLMs desarrollados en China son impresionantes en lo técnico, útiles en lo práctico y expansivos en su visión. Pero no están exentos de debate:

¿Qué significa “IA ética” bajo un sistema donde el control estatal prima sobre la libertad de expresión?

¿Hasta qué punto pueden ser adoptados globalmente sin sesgos ideológicos?

A pesar de estas preguntas, no se puede ignorar el peso que China ya tiene en el tablero global de la IA. Su producción masiva de modelos, su capacidad de implementación en tiempo récord y su apertura hacia el mundo no occidental podrían redefinir el equilibrio global de la inteligencia artificial en esta década.

5.3.5. Comparativa visual: LLMs de China, EE.UU. y Europa en educación

A medida que la inteligencia artificial entra con fuerza en las aulas, universidades y sistemas de formación, resulta clave entender qué ofrece cada región, qué prioridades tiene y cómo se reflejan en sus modelos de lenguaje masivo.

 ¿Qué prioriza cada región en educación con IA?

 Estados Unidos:

- ♦ Innovación rápida
- ♦ Modelos como ChatGPT son populares por su adaptabilidad, utilizados para redacción de ensayos, preparación de exámenes, explicaciones personalizadas y apoyo en programación.

 Europa:

- ♦ IA con propósito social
- ♦ Modelos como BLOOM y Aleph Alpha priorizan educación multilingüe, soberanía digital y ética. Especialmente útiles en universidades públicas, ONGs y proyectos comunitarios.

 China:

- ♦ Universalización educativa
- ♦ Modelos como SparkDesk y DeepSeek se orientan a ampliar el acceso a educación de calidad en zonas remotas, ofreciendo tutoría automatizada y contenidos adaptados a currículos oficiales.

 Tabla 5.13 IA en educación según su origen

Región	Modelos clave	Enfoque educativo principal	Fortalezas didácticas	Riesgos o limitaciones
 EE.UU.	ChatGPT (OpenAI), Claude (Anthropic), Gemini (Google)	Versatilidad, creatividad, tutoría personalizada	Multimodalidad, asistentes para redacción, codificación y estudio autónomo	Riesgos de sesgos culturales, privacidad con plataformas comerciales
 Europa	Aleph Alpha, Mistral, BLOOM, Noosa	Inclusión, ética, transparencia, educación pública	Multilingüismo, explicabilidad, cumplimiento de normativas como GDPR	Menor escalabilidad y velocidad, limitada integración comercial
 China	DeepSeek-V3, ERNIE 4.0, SparkDesk, Tongyi	Igualdad de acceso, enfoque práctico, escala nacional	Educación en zonas rurales, asistentes por voz, rendimiento en matemáticas	Control estatal del contenido, restricciones ideológicas, filtrado automático

 Cada región construye su IA desde sus valores.

 La educación del futuro no dependerá solo de tener la mejor tecnología, sino de elegir aquella que resuene con los principios y necesidades de cada comunidad educativa.

 ¿Qué modelo elegir según tus prioridades?

Tabla 5.14

Si buscas...	Elige modelos de...	¿Por qué?
Creatividad, versatilidad y herramientas listas	EE.UU.	Ecosistema comercial desarrollado y tecnologías maduras
Transparencia, ética y cumplimiento normativo	Europa	Diseñados desde una base legal, social y educativa equilibrada

Escalabilidad, educación en gran escala	China	Implementación masiva y adaptabilidad en contextos institucionales
---	-------	--

5.4. Panorama global y retos

5.4.1. La carrera por el dominio de la IA

La inteligencia artificial generativa se ha convertido en el nuevo campo de batalla tecnológico y geopolítico del siglo XXI. Ya no se trata solo de innovación científica, sino de poder, soberanía y hegemonía cultural.

Dominar los LLMs no es solo una cuestión de quién puede generar mejores respuestas, sino de quién define las reglas del futuro digital.



Figura 5.24 La guerra por hegemonía en IA: EU vs China

Mientras que Estados Unidos lidera en capacidad empresarial y Europa en ética normativa, China emerge como un jugador con ambición total: desarrollo acelerado, modelos propios, y un ecosistema cerrado pero extremadamente eficiente.

5.4.2. Costo de entrenar un LLM: la otra barrera de entrada

Detrás de cada modelo de IA exitoso, hay una inversión masiva que pocos pueden asumir. Entrenar un modelo como GPT-4 puede costar más de 100 millones de dólares, y eso no incluye mantenimiento, actualizaciones ni adaptación continua.

¿Qué hace que un LLM sea tan costoso?

-  Infraestructura computacional: miles de GPUs trabajando en paralelo (a menudo A100 o H100 de NVIDIA).

- 📦 Datos de entrenamiento: billones de palabras, código, imágenes y estructuras semánticas.
- ⚡ Consumo energético: entrenar un LLM puede consumir tanta energía como cientos de hogares durante meses.
- 🧠 Talento técnico: equipos multidisciplinares altamente especializados en IA, matemáticas, lingüística y ética.

💡 Esto significa que el desarrollo de IA generativa no está al alcance de todos los países ni todas las universidades. Hay un muro de entrada que separa a los desarrolladores de IA en dos mundos: los que pueden crearla... y los que solo pueden usarla.

🏛️ 5.4.3. ¿Puede un país controlar el futuro de la IA?

La respuesta corta: sí, pero no solo.

La respuesta larga: ningún país puede controlar la IA globalmente, pero sí puede liderar su influencia dentro de su territorio, idioma, cultura y política.

- Estados Unidos domina el mercado y la narrativa mediática.
- China controla su ecosistema con soberanía absoluta.
- Europa regula con firmeza para proteger derechos digitales.

Pero también emergen otros actores:

- India y su apuesta por modelos en hindi, bengalí y otras lenguas regionales.
- Emiratos Árabes con Falcon LLM, modelo open-source con ambición global.
- América Latina, África y el Sudeste Asiático explorando modelos propios o adaptados.

🕸️ El control ya no es solo sobre los modelos, sino sobre los datos, las plataformas y la infraestructura que los aloja.

🗺️ 5.4.4. “Conquistar la cima del liderazgo tecnológico”: visión geopolítica

La frase, citada en documentos estratégicos de Estados Unidos, China y Emiratos, resume el verdadero objetivo: no solo desarrollar IA, sino alcanzar la cima tecnológica antes que otros.

La IA es ahora:

- Una herramienta de defensa nacional.
- Un activo de influencia cultural.
- Un componente clave del PIB digital.

China como caso emblemático:

🇨🇳 Estrategia China:

- Desarrollo nacional autónomo (sin depender de chips o plataformas de EE. UU.).
- Inversión masiva en modelos propios (ERNIE, Tongyi, DeepSeek).
- Integración total con WeChat, Alibaba Cloud, sistemas escolares y cámaras urbanas.
- Diplomacia tecnológica con países del sur global (BRICS, África, América Latina).

 *China no quiere solo competir: quiere rediseñar el terreno de juego.*

5.4.5. ¿Y ahora qué?

La inteligencia artificial generativa no es neutra. Refleja las visiones del mundo de quienes la diseñan, financian y regulan.

La gran pregunta no es qué puede hacer la IA, sino quién tiene el poder para decidir cómo debe hacerlo.

Y en ese contexto, cada docente, institución, país o comunidad debe plantearse:

- ¿Qué modelos queremos adoptar?
- ¿Qué valores queremos preservar?
- ¿A quién estamos entregando nuestra autonomía digital?

El reto no es solo tecnológico. Es político, educativo y cultural.

5.5 Conclusiones

Este capítulo nos ha permitido abrir la caja negra de la inteligencia artificial generativa y comprender cómo funcionan realmente los modelos de lenguaje de gran escala (LLMs), esos motores invisibles que hoy están transformando la educación, el trabajo, la investigación y la creatividad.

 La IA generativa no solo interpreta: **crea**. Puede redactar textos, generar imágenes, producir audio, escribir código o elaborar presentaciones en segundos.

 Los LLMs —como GPT-4, Claude o Gemini— no tienen conciencia, pero son capaces de **reconocer patrones lingüísticos** con precisión sorprendente, generando contenido que parece humano.

 Detrás de su funcionamiento están tres pilares clave:

- La arquitectura **Transformer**, que revolucionó el procesamiento del lenguaje natural.
- Los **tokens**, unidades mínimas de lenguaje computacional que permiten construir respuestas complejas.
- Las **ventanas de contexto**, que determinan cuánta información pueden “recordar” para mantener coherencia en una conversación o documento largo.

 El mapa global de la IA generativa es también un escenario de disputa **geopolítica y estratégica**:

- Estados Unidos lidera con OpenAI, Google y Anthropic.
- China avanza con modelos como ERNIE y Qwen.
- Europa apuesta por soberanía tecnológica y ética con iniciativas como LLaMA y BLOOM.

 Entender esta tecnología no es solo un asunto técnico, sino cultural y estratégico. La IA generativa ya no es una curiosidad: es infraestructura clave del conocimiento contemporáneo.

¿Qué sigue en nuestro viaje?

En el próximo capítulo nos adentraremos en el **impacto profundo de la inteligencia artificial en la educación**.

Exploraremos:

- ¿Cómo está cambiando el rol del docente y del estudiante?
- ¿Qué formas de aprendizaje emergen con el uso de IA?
- ¿Cómo diseñar entornos educativos que aprovechen la IA sin perder el factor humano?

Prepárate para descubrir cómo la educación puede reinventarse con el poder de la inteligencia artificial —y qué papel jugarás tú en ese cambio.

CHATGPT EN ACCIÓN

Aprende haciendo con ChatGPT

OBJETIVO:

Experimentar de forma práctica con las capacidades, limitaciones y usos estratégicos de los modelos de lenguaje de gran escala, comprendiendo cómo funcionan y cómo pueden aplicarse en contextos educativos, creativos y profesionales.

HERRAMIENTAS A UTILIZAR:

- ChatGPT (versión gratuita)
- Acceso a internet para exploración de plataformas (opcional)

ACTIVIDAD PRÁCTICA N.º 1

Simula el funcionamiento de un LLM

Actividad:

Pide a ChatGPT que te explique cómo predice una respuesta, paso a paso, y simula cómo él "piensa" al completar una frase o pregunta compleja.

Prompt sugerido para ChatGPT:

"Explícame paso a paso cómo un modelo como ChatGPT predice la siguiente palabra en una frase. Simula el proceso usando esta oración: 'La educación del futuro será...'"

Propósito:

Comprender cómo funciona internamente un modelo generativo y qué significa "predecir tokens".

ACTIVIDAD PRÁCTICA N.º 2

Explora tipos de contenido que puede generar la IA

Actividad:

Interactúa con ChatGPT para generar un ejemplo de cada tipo de contenido: texto, imagen (descripción), código, audio (simulado) y video (guionizado).

Prompt sugerido para ChatGPT:

"Quiero que actúes como una IA generativa multimodal. Crea para mí: (1) un poema breve, (2) una descripción para generar una imagen de portada de ebook, (3) un

fragmento de código en Python para una calculadora, (4) una idea de guion para un video educativo de 1 minuto y (5) un concepto para un audio de meditación."

 Sugerencia:

Comparte los resultados con colegas para debatir sobre la creatividad y utilidad de cada tipo de contenido.

 ACTIVIDAD PRÁCTICA N.º 3

Compara modelos de IA actuales

 Actividad:

Haz que ChatGPT actúe como un asesor tecnológico y te ayude a comparar dos modelos de IA generativa para un uso educativo.

Prompt sugerido para ChatGPT:

"Soy docente universitario. ¿Cuál modelo de IA generativa me recomendarías entre GPT-4 y Claude 3 para crear contenidos educativos interactivos? Compáralos según comprensión de contexto, costo, facilidad de uso y privacidad."

 Utilidad:

Desarrollar criterio para elegir herramientas de IA según necesidades específicas.

 ACTIVIDAD PRÁCTICA N.º 4

Diseña una aplicación educativa con IA

 Actividad:

Usa ChatGPT para ayudarte a imaginar una app educativa que integre IA mediante una API. Define su objetivo, público, funciones y cómo usaría un modelo de lenguaje.

Prompt sugerido para ChatGPT:

"Ayúdame a diseñar una app educativa para secundaria que use una API de IA generativa. Quiero que incluya tutoría automática, resúmenes y test de autoevaluación. ¿Cómo funcionaría y qué ventajas ofrecería?"

 Propósito:

Conectar el conocimiento técnico con una propuesta realista de aplicación en el aula o en entornos formativos.

ACTIVIDAD PRÁCTICA N.º 5

Resumen reflexivo del capítulo 5

Actividad:

Crea un resumen que integre conceptos como IA generativa, LLMs, tokens, ventanas de contexto, APIs y tipos de modelos (abiertos vs cerrados).

Prompt sugerido para ChatGPT:

"Ayúdame a redactar un resumen de 250 palabras del capítulo 5 del libro. Quiero que incluyas qué es la IA generativa, cómo funcionan los modelos LLM, qué tipos de contenido pueden generar, y una reflexión sobre los riesgos y ventajas de los modelos cerrados y abiertos."

Propósito:

Sintetizar aprendizajes complejos en una visión crítica y aplicada, orientada al uso consciente de la IA.

HERRAMIENTAS IA: Tu nuevo kit de superpoderes

 Herramienta recomendada: **Elicit**

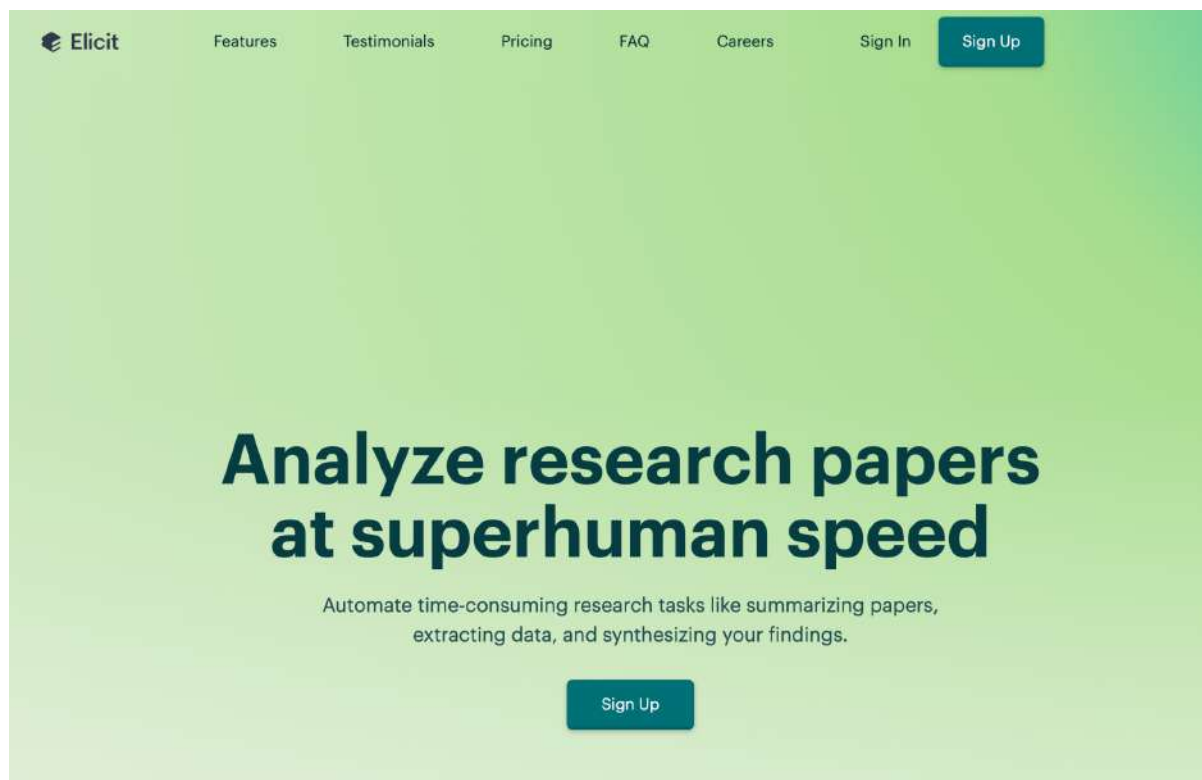


Figura 5.24 Elicit <https://elicit.com/welcome>

¿Qué es?

Elicit es una herramienta de inteligencia artificial diseñada para facilitar tareas de investigación académica. Utiliza modelos de lenguaje para ayudar a buscar, filtrar, resumir y organizar artículos científicos, convirtiéndose en un asistente ideal para investigadores, docentes y estudiantes avanzados. Se basa en fuentes como Semantic Scholar y permite automatizar procesos clave en revisiones de literatura.

¿Para qué sirve en educación?

Es especialmente útil en contextos universitarios o de investigación. Elicit permite ahorrar tiempo al buscar papers relevantes, extraer ideas clave, comparar hallazgos y organizar referencias. También es una excelente

herramienta para enseñar metodología de investigación, lectura crítica y redacción académica asistida.

✓ Ventajas clave

- Búsqueda inteligente de artículos científicos con IA.
- Extracción automática de resúmenes, variables y métodos de estudio.
- Comparación entre papers por criterio o pregunta de investigación.
- Organización de resultados en tablas editables exportables.
- Interfaz intuitiva y sin curva técnica elevada.
- Ayuda a desarrollar pensamiento crítico y habilidades de síntesis.

 Accede aquí: <https://elicit.org>

💰 Modelo de precios:

- Plan gratuito: acceso completo a la mayoría de las funciones.
- Planes premium (en desarrollo): podrían incluir más capacidad de procesamiento o funcionalidades para equipos de investigación.

Pick a plan that's right for you

Monthly ☒ Annually

Basic	Plus	Pro	Team	Enterprise
For casual exploration	For deeper research	For systematic reviews	For collaboration	For companies and schools
Free	\$10 per month, billed annually \$120 per year	\$42 per month, billed annually \$499 per year	\$65 per user, per month billed annually \$780	Custom
Everything in Basic and...	Everything in Plus and...	Everything in Pro and...	Everything in Team and...	Everything in Enterprise and...
✓ Automated research reports ✓ Unlimited search across more than 125 million papers ✓ Unlimited summaries of 4 papers at once ✓ Unlimited chat with 4 papers at once ✓ Extract data from 20 papers per month	✓ Unlimited chat with 8 full-text papers at once ✓ Unlimited summaries of 8 papers at once ✓ Export to RIS, CSV, Bib ✓ Extract data from 600 papers per year ✓ Add 5 columns to your tables at a time	✓ Extract data from 2,400 papers per year ✓ Extract data from tables inside papers ✓ Dedicated workflows for systematic reviews ✓ Add 20 columns to your tables at a time ✓ Binary & multi-select fields for paper coding	✓ Extract data from up to 3,600 papers per user per year, pooled across your team ✓ Add 30 columns to your tables at a time ✓ Live-editing for Notebooks and Systematic Reviews ✓ Admin panel with usage tracking and seat management	✓ Custom usage limits ✓ Volume discounts ✓ Training and onboarding ✓ Data configuration ✓ Co-branding of Notebooks and assets ✓ Custom workflow development

Figura 5.25 Planes de la Herramienta Elicit

GUÍA DE ESTUDIO

REFERENCIAS

Anthropic. (2023). *Constitutional AI: Training language models to follow principles*.

<https://www.anthropic.com/index/constitutional-ai>

Caracol Radio. (2024, septiembre 1). *Los 10 países que más usan ChatGPT: solo uno latino y no es Colombia*.

<https://caracol.com.co/2024/09/01/los-10-paises-que-mas-usan-chatgpt-solo-uno-latino-y-no-es-colombia/>

DeepSeek AI. (2024). *DeepSeek v3 Technical Overview*. <https://www.deepseek.com/>

Email Vendor Selection. (2024). *ChatGPT estadísticas y demografía de usuarios*.

<https://www.emailvendorselection.com/es/chatgpt-estadisticas/>

Exploding Topics. (2024, diciembre). *ChatGPT User Statistics: How Many People Use ChatGPT in 2024?* <https://explodingtopics.com/blog/chatgpt-users>

Google DeepMind. (2023). *Introducing Gemini: A new era for AI at Google*.

<https://deepmind.google/technologies/gemini/>

Hugging Face. (2023). *Open LLM Leaderboard*.

https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard

Martínez, A. (2023, febrero 27). *ChatGPT logró en solo cinco días obtener un millón de usuarios*. *El Economista*.

<https://www.eleconomista.com.mx/tecnologia/ChatGPT-logro-en-solo-cinco-dias-obtener-un-millon-de-usuarios-20230227-0020.html>

Meta AI. (2023). *LLaMA: Open Foundation Language Models*. <https://ai.meta.com/llama/>

Microsoft. (2023). *Azure OpenAI Service*.

<https://azure.microsoft.com/en-us/products/cognitive-services/openai-service/>

OpenAI. (2023). *Introducing GPT-4*. <https://openai.com/research/gpt-4>

OpenAI. (2023). *Pricing*. <https://openai.com/pricing>

Sejnowski, T. J. (2023). *Reflections on "Attention Is All You Need"*. NeurIPS Talks. <https://neurips.cc/>

Statista. (2024). *Principales países usuarios de ChatGPT en Latinoamérica*. <https://es.statista.com/estadisticas/1382643/principales-paises-usuarios-chatgpt-en-latinoamerica/>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention Is All You Need*. *Advances in Neural Information Processing Systems*. https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Vincent, J. (2023, October 17). *New US restrictions block AI chip sales to China, targeting Nvidia*. *The Verge*. <https://www.theverge.com/>

Zhang, X., & Lin, K. (2024). *China's Push for AI Supremacy: Models, Policies, and Platforms*. *AI Policy Review*, 12(2), 115–136.