

Modelos de Lenguaje

Propietarios y Open Source

Introducción a los Grandes Modelos de Lenguaje

Aprende qué son las LLMs como
ChatGPT y su impacto en nuestra
sociedad

[Seguir Leyendo](#)

@jbagnato | www.aprendemachinelearning.com



Los modelos de lenguaje son sistemas de inteligencia artificial diseñados para comprender, interpretar y generar lenguaje humano de manera natural y coherente.

Fundamentos Técnicos



¿Qué son los LLM?

Los Large Language Models (LLM) son sistemas de IA entrenados con enormes cantidades de datos textuales para comprender y generar lenguaje humano. Utilizan redes neuronales profundas y pueden realizar diversas tareas lingüísticas.



Arquitectura Transformer

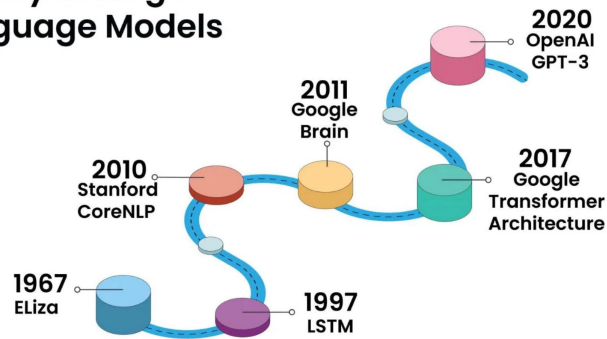
Los LLM modernos utilizan la arquitectura Transformer, que emplea mecanismos de atención para procesar secuencias de texto en paralelo, permitiendo capturar relaciones complejas entre palabras independientemente de su distancia.



Tokens y Procesamiento

El texto se divide en unidades llamadas tokens (palabras o subpalabras). Los modelos tienen límites de contexto que determinan cuántos tokens pueden procesar simultáneamente, desde miles hasta millones en los más avanzados.

History of Large Language Models



Proceso de Entrenamiento

Pre-entrenamiento: El modelo es entrenado supervisado en enormes corpus de texto donde el modelo predice palabras en contexto.

Fine-tuning: Ajuste con datos específicos y retroalimentación humana para mejorar capacidades y alineación con valores.

Clasificación de Modelos de Lenguaje

Por Arquitectura

Los modelos se clasifican según su diseño arquitectónico:

Transformer:
Arquitectura dominante basada en mecanismos de atención (GPT, BERT)

Recurrentes (RNN):
Procesan secuencias en orden (modelos más antiguos)

Híbridos: Combinan diferentes arquitecturas para tareas específicas

Por Tamaño

El tamaño se mide en número de parámetros:

Pequeños: Menos de 10B parámetros (GPT-3.5, Mistral-7B)

Medianos: Entre 10B y 70B parámetros (Claude 3 Sonnet, LLaMA-70B)

Grandes: Más de 70B parámetros (GPT-4, Claude 4 Opus)

Por Licencia

Según su disponibilidad y condiciones de uso:

Propietarios:
Acceso controlado, APIs de pago (OpenAI, Anthropic, Google)

Open Source: Código y pesos disponibles (LLaMA, Mistral, Falcon)

Híbridos: Modelos base abiertos con versiones comerciales optimizadas

Por Capacidades

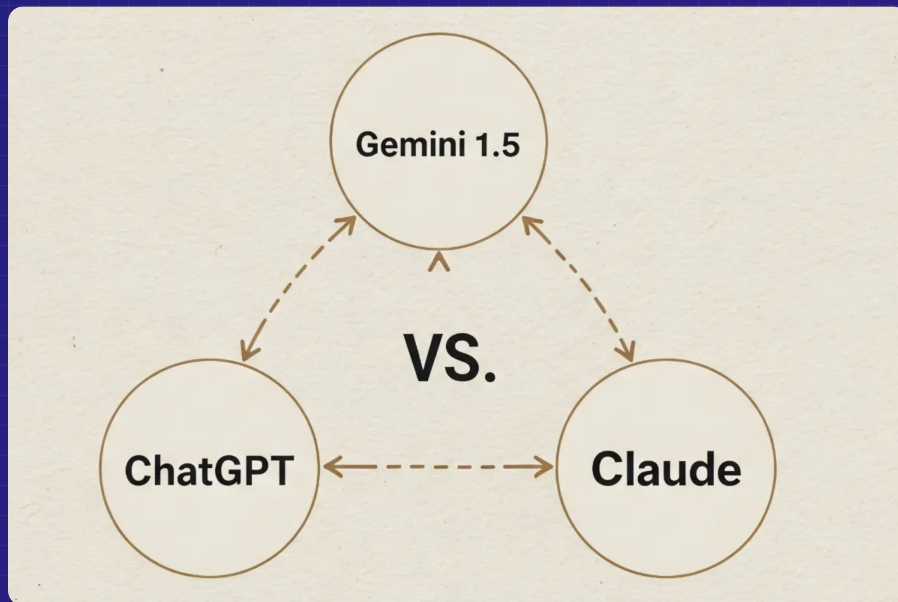
Según las modalidades que pueden procesar:

Solo texto: Procesan únicamente entrada y salida textual

Multimodales: Manejan texto, imágenes, audio o video (GPT-4o, Gemini)

Especializados:
Optimizados para dominios específicos (código, medicina)

Modelos de Lenguaje Proprietarios



OpenAI

GPT-4o / o3

- ✓ Capacidades multimodales (texto, imagen, audio)

Anthropic

Claude 4 Sonnet / Opus

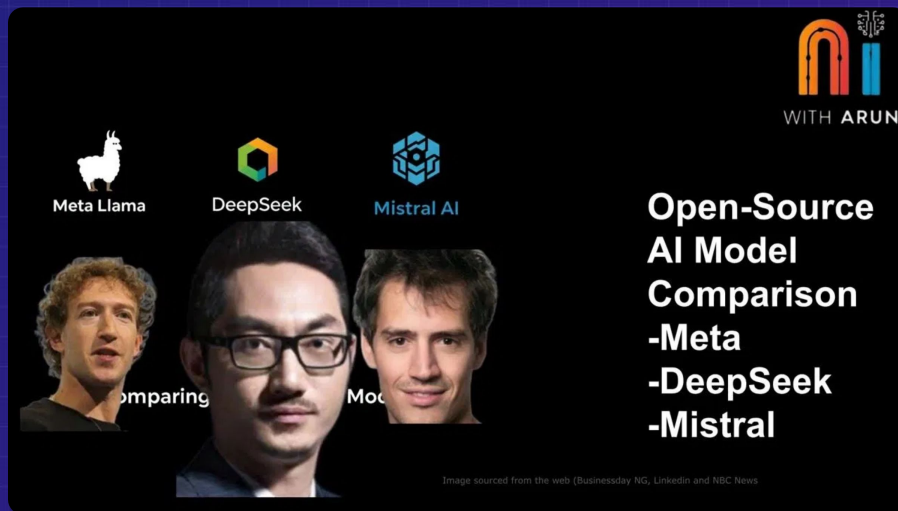
- ✓ Excelente para generación de páginas web

Google

Gemini 2.5 Pro

- ✓ Integración con herramientas de Google

Modelos de Lenguaje Open Source



Meta

Familia LLaMA

- ✓ Versiones de 8B a 70B parámetros
- ✓ Licencia permisiva para uso comercial
- ✓ Variantes especializadas (CodeLlama)

Mistral AI

Mistral

- ✓ Modelo francés de alto rendimiento
- ✓ Excelente para programación en Python
- ✓ Versiones de 7B a 8x7B parámetros

DeepSeek

DeepSeek Coder / R1

- ✓ Modelo chino especializado en razonamiento
- ✓ Modo DeepThink (Chain of Thoughts)
- ✓ Versión especializada para programación

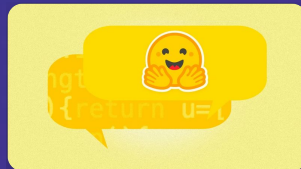
Plataformas y Servicios



Groq

Plataforma que hostea modelos de inteligencia artificial open source con hardware especializado.

- ⚡ Velocidad de respuesta ultrarrápida
- 💻 Hardware LPU (Language Processing Unit)
- ✅ Soporte para múltiples modelos open source



Hugging Face

Principal repositorio y comunidad de modelos de lenguaje open source.

- 🗄 Miles de modelos disponibles para descarga
- </> Herramientas para implementación y fine-tuning
- 🧪 Sección "Spaces" para probar modelos



LMarena.ai

Plataforma de evaluación y comparación de modelos de lenguaje.

- 📊 Benchmarks comparativos entre modelos
- 🏆 Rankings actualizados de rendimiento
- ⚖ Evaluación objetiva de capacidades

Otras Plataformas

🔗 **LangChain:** Framework para desarrollo de aplicaciones basadas en LLM

Recursos de Desarrollo

🔗 **GitHub:** Repositorios de código y modelos

Comparativa de Modelos

Modelo	Tipo	Rendimiento	Multimodalidad	Contexto	Costo	Caso de uso ideal
GPT-4o	Propietario	★★★★★	Alta	128K	\$0-200/mes	Uso general, multimodal
Claude 4 Opus	Propietario	★★★★★	Media	200K	\$20-200/mes	Investigación, codificación
Gemini 2.5 Pro	Propietario	★★★★☆	Alta	1M	\$19.99-249.99/mes	Análisis de documentos extensos
LLaMA 3 (70B)	Open Source	★★★★☆	Baja	32K	Gratuito	Implementación local, personalización
Mistral Large	Open Source	★★★★☆	Baja	32K	Gratuito	Programación, eficiencia
DeepSeek	Open Source	★★★★☆	Baja	16K	Gratuito	Razonamiento, Chain of Thought

i Consideraciones importantes:

La elección del modelo ideal depende de factores como el caso de uso específico, los recursos disponibles, los requisitos de privacidad y el presupuesto. Los modelos propietarios ofrecen mayor rendimiento y capacidades multimodales, mientras que los open source brindan flexibilidad, personalización y control sobre los datos.

Aplicaciones Prácticas



Generación de Texto Creativo

Los modelos de lenguaje destacan en la creación de contenido original y creativo.

- ✓ Redacción de artículos, blogs y contenido web
- ✓ Creación de guiones, historias y narrativas
- ✓ Generación de ideas y conceptos creativos



Programación y Desarrollo

Asistencia en tareas de programación y desarrollo de software.

- ✓ Generación y depuración de código
- ✓ Explicación de conceptos técnicos
- ✓ Optimización y refactorización de código



Investigación y Análisis

Apoyo en procesos de investigación y análisis de información.

- ✓ Resumen y síntesis de documentos extensos
- ✓ Análisis de datos y generación de informes
- ✓ Respuesta a preguntas complejas con razonamiento



Educación

Herramientas para mejorar procesos educativos y de aprendizaje.

- ✓ Tutorías personalizadas y resolución de dudas
- ✓ Creación de materiales didácticos
- ✓ Simulación de debates y discusiones

Tendencias y Futuro

Modelos Multimodales

La integración de múltiples modalidades (texto, imagen, audio, video) en un solo modelo está transformando las capacidades de los LLM. Los modelos como GPT-4o y Gemini 2.5 Pro pueden procesar y generar contenido en diferentes formatos, ampliando significativamente sus aplicaciones prácticas.

Razonamiento en Cadena

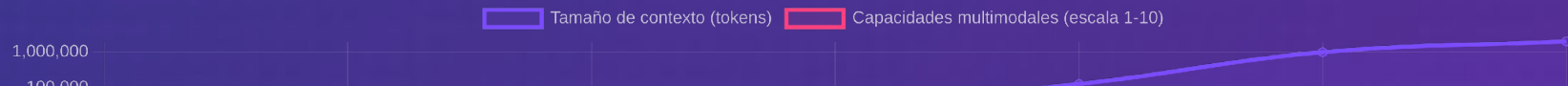
Los modelos están evolucionando hacia un razonamiento más sofisticado mediante técnicas como Chain of Thought (CoT). Modelos como o3 y DeepSeek R1 pueden descomponer problemas complejos en pasos lógicos, mejorando significativamente su capacidad para resolver tareas que requieren razonamiento avanzado.

Ventanas de Contexto Ampliadas

La capacidad de procesar contextos cada vez más extensos está en constante evolución. Desde los 2K tokens iniciales hasta el millón de tokens de Gemini 2.5 Pro, esta tendencia permite analizar documentos completos, mantener conversaciones extensas y comprender información a gran escala.

Reducción de Costos

Los costos de entrenamiento e inferencia están disminuyendo, haciendo que la IA avanzada sea más accesible. La optimización de modelos, hardware especializado como las LPU de Groq y la competencia entre proveedores están democratizando el acceso a estas tecnologías.



Recomendaciones y Conclusiones

Cómo elegir el modelo adecuado

- 🎯 **Define tu objetivo:** Identifica claramente qué tarea necesitas realizar (generación de texto, programación, análisis de datos, etc.)
- ⚖️ **Evalúa tus recursos:** Considera tu presupuesto, infraestructura técnica y requisitos de privacidad
- 🔍 **Prueba diferentes modelos:** Experimenta con varios modelos para encontrar el que mejor se adapte a tus necesidades específicas
- 🛡️ **Considera la ética:** Evalúa las implicaciones éticas y de privacidad del uso de cada modelo

Actividades propuestas

- ☞ **Comparación práctica:** Probar los distintos modelos de lenguaje vistos en clase y sacar conclusiones sobre sus fortalezas y debilidades
- 🔍 **Investigación:** Buscar otros LLM que no hayamos visto en clase y analizar sus características
- 🧪 **Experimentación:** Probar modelos Open Source de HuggingFace en la sección "Spaces"

🔗 Recursos adicionales

🔗 GitHub del Profesor: github.com/esteban-calabria/Intro-la/

🔗 Hugging Face: huggingface.co