

1.3 ¿Qué son los LLM?

Los **grandes modelos de lenguaje** (LLM, por las siglas de su denominación en inglés, Large Language Models) son algoritmos “entrenados” con grandes volúmenes de documentos humanos para generar textos como respuesta a peticiones (*prompts*) específicas. Se trata de modelos estadísticos capaces de predecir la siguiente palabra o grupo de palabras gracias a un entrenamiento “afinado” por medio de supervisores humanos.

Mucha gente identifica la Inteligencia Artificial con los Grandes Modelos de Lenguaje, como las sucesivas encarnaciones de ChatGPT. Esta identificación, sin embargo, no es del todo exacta, y puede llevar a confusiones. **No todas las herramientas de Inteligencia Artificial son iguales**, ni en su configuración, ni en su entrenamiento ni en su propósito. La aparición de modelos integrados de inteligencia artificial generativa capaces de responder al lenguaje "natural" ha contribuido a incrementar la confusión.

En el siguiente vídeo podéis ver una introducción sencilla a algunos conceptos básicos relacionados con la historia y el funcionamiento de los LLM:

<https://www.youtube.com/embed/Sz4qacFBHLk>

También podéis encontrar un análisis **más técnico y detallado** (y muy didáctico) en [este módulo](#) creado para CATEDU por Luis Hueso Ibañez y Pedro López Savirón como parte del curso *La IA en educación. Una aproximación práctica*.

A continuación incluimos un **glosario** con los términos que nos pueden resultar más útiles para un enfoque centrado en las materias de perfil sociolingüístico:

IA (Inteligencia Artificial)	Conjunto de tecnologías que permiten a las máquinas realizar tareas que normalmente requieren inteligencia humana, como comprender texto, reconocer imágenes o tomar decisiones simples.
IA generativa	Tipo de IA capaz de crear contenido nuevo (texto, imágenes, audio, código) a partir de patrones aprendidos en grandes conjuntos de datos.

LLM (Large Language Model)	Modelo de IA entrenado con enormes cantidades de texto para comprender y generar lenguaje natural. Es la base de muchos asistentes conversacionales actuales.
GPT (Generative Pre-trained Transformer)	Tipo de modelo de lenguaje basado en la arquitectura transformer que ha sido preentrenado con grandes cantidades de texto para generar respuestas coherentes.
Procesamiento del lenguaje natural (PLN / NLP)	Campo de la IA que se ocupa de que los ordenadores comprendan, interpreten y produzcan lenguaje humano.
Aprendizaje automático (Machine Learning)	Técnica mediante la cual los sistemas informáticos aprenden patrones a partir de datos en lugar de ser programados con reglas explícitas.
Aprendizaje profundo (Deep Learning)	Subcampo del aprendizaje automático que utiliza redes neuronales con múltiples capas para analizar grandes cantidades de datos.
Redes neuronales	Modelos matemáticos inspirados en el cerebro humano formados por capas de nodos conectados que procesan información y detectan patrones.
Aprendizaje supervisado	Método de entrenamiento en el que el modelo aprende a partir de datos previamente etiquetados por humanos.
Token	Unidad mínima de texto que un modelo de lenguaje procesa (palabra, parte de una palabra o signo de puntuación).
Chatbot (asistente conversacional)	Programa que permite interactuar con la IA mediante conversación en lenguaje natural.
Prompt	Instrucción o pregunta que se introduce en un sistema de IA para generar una respuesta o contenido.
Ingeniería de prompts	Conjunto de técnicas para formular instrucciones eficaces y obtener mejores resultados de un sistema de IA.
Agente de IA	Sistema de IA que puede planificar acciones, tomar decisiones y utilizar herramientas externas para cumplir un objetivo (algunos sistemas, por ejemplo ChatGPT, cuentan con un Agent Mode que permite encadenar tareas y actuar con mayor autonomía).
Deep Research	Capacidad de algunos sistemas de IA para realizar investigaciones complejas, consultando múltiples fuentes y generando informes estructurados.
AGI (Inteligencia Artificial General)	Hipotético tipo de IA capaz de realizar cualquier tarea intelectual humana con flexibilidad y autonomía. Actualmente no existe.
Alucinación	Situación en la que un sistema de IA genera información falsa o «inventada» presentándola como si fuera correcta.

Sesgo (Bias)	Distorsión en los resultados de la IA causada por datos de entrenamiento incompletos o prejuicios presentes en los datos.
Alineamiento	Área de investigación que busca que los sistemas de IA actúen de acuerdo con valores humanos, normas sociales y objetivos seguros.
Supervisión humana (Human-in-the-loop)	Principio según el cual las decisiones o resultados de la IA deben ser revisados o validados por personas.

Revision #6

Created 2026-03-10 18:15:51 CET by Miguel Serrano Larraz

Updated 2026-03-16 13:51:21 CET by Miguel Serrano Larraz