3.1 Introducción y modelos fundacionales

Introducción

Una vez que tenemos diseñada nuestra Situación de Aprendizaje, quizá necesitemos elaborar diferentes materiales o recursos didácticos a emplear durante el desarrollo de la misma. Para ello veremos qué utilidades nos pueden ofrecer algunos sistemas de inteligencia artificial generativa.

> Recuerda que el **contenido que le proporciones a la IA** ha de ser de **autoría propia o** disponer de la **licencia** para poder utilizar ese contenido.

> Te recomendamos que para probar herramientas que requieran registro, te abras una nueva **cuenta de correo electrónico específica** para las herramientas o aplicaciones, **así evitarás spam innecesario** en tu cuenta principal. También puedes probar con cuentas de correo temporal como tempmail.

> Es posible que las funciones, características o precio de las herramientas que se presentan en el siguiente capítulo cambien en un futuro. Revisión en marzo de 2026.

Modelos fundacionales: base de los recursos de texto, audio, imagen y vídeo

Copyright 2026

Cuando hablamos de generación automática de texto, transcripción de audio, creación de imágenes o análisis de vídeo con IA, en realidad estamos hablando de una misma idea tecnológica de fondo: los **modelos fundacionales**.

Un modelo fundacional es un modelo de inteligencia artificial entrenado con enormes cantidades de datos y diseñado para servir como base común sobre la que se construyen múltiples aplicaciones. No nace para hacer una única tarea concreta, sino para adaptarse a muchas: redactar textos, resumir documentos, traducir idiomas, describir imágenes, generar código, sintetizar voz o analizar contenido multimedia.

Empresas como OpenAI, Google DeepMind, Anthropic, Meta o Mistral AI han desarrollado modelos fundacionales que actúan como motores generales de inteligencia artificial. A partir de ellos se construyen asistentes conversacionales, generadores de presentaciones, sistemas de análisis documental, herramientas de edición multimedia o plataformas educativas.

La característica principal de estos modelos es su carácter **generalista**. A diferencia de los modelos tradicionales, que se entrenaban para una única tarea (por ejemplo, detectar spam o clasificar imágenes médicas), los modelos fundacionales se entrenan primero de forma masiva y después se adaptan mediante prompting, ajuste fino (fine-tuning) o integración con bases de datos externas (RAG). Esto permite reutilizar el mismo modelo para múltiples contextos.

En el ámbito de los recursos educativos y profesionales, esta base común explica por qué hoy podemos trabajar con:

- Generación y transformación de textos (resúmenes, esquemas, traducciones, rúbricas).
- Transcripción automática y síntesis de voz.
- Generación y análisis de imágenes.
- Creación o comprensión de vídeo.
- Sistemas multimodales que combinan texto, imagen y audio en una misma interacción.

Muchos de estos modelos están construidos sobre arquitecturas tipo Transformer y pueden operar en una o varias modalidades. Algunos son exclusivamente de texto (LLMs), otros están especializados en visión o audio, y los más recientes son multimodales, capaces de integrar diferentes tipos de información en un único sistema.

Además, los modelos fundacionales pueden ser:

- **Propietarios**, accesibles mediante API.
- **Abiertos**, con pesos descargables y ejecutables en local (algo especialmente relevante en entornos educativos donde se busca privacidad, control y experimentación técnica).

En definitiva, los modelos fundacionales son el “motor” que hace posible todos los recursos de generación y análisis de texto, audio, imagen y vídeo que veremos en este capítulo. Entender su

Copyright 2026

naturaleza, sus límites y sus posibilidades es clave para utilizar estas herramientas de manera crítica, técnica y responsable.

Modalidad / Uso	Modelo / Ejemplo	Descripción breve	Aplicación típica
Texto / Lenguaje	GPT (OpenAI)	Modelo de lenguaje preentrenado capaz de generación y comprensión de texto.	Bots conversacionales, resúmenes y traducción de textos.
	Claude (Anthropic)	LLM orientado a respuestas coherentes y seguras.	Asistencia en escritura y análisis de texto.
Imagen	Stable Diffusion	Modelo generador de imágenes a partir de descripciones en texto.	Creación de ilustraciones y diseños visuales.
	DALL-E	Generación de imágenes creativas desde instrucciones textuales.	Arte digital y gráficos personalizados.
Audio / Voz (Reconocimiento – Speech-to-Text)	Whisper (OpenAI)	Modelo para transcripción automática de voz a texto en múltiples idiomas.	Transcripción de grabaciones y clases.
	Modelos de reconocimiento y traducción de voz (varios)	Modelos entrenados para convertir audio hablado en texto y/o traducir voz.	Subtítulos automáticos y traducción de audio.
Audio / Voz (Síntesis – Text-to-Speech)	Modelos TTS comerciales	Modelos que generan voz natural desde texto.	Narración de contenidos y asistentes de voz.
Vídeo (Generación y Multimodal audiovisual)	Sora (OpenAI)	Modelo de IA que genera vídeos realistas a partir de descripciones textuales, incluyendo audio.	Producción de clips de vídeo con sonido integrado.
	Veo 3 (Google DeepMind)	Generador de vídeos con audio sincronizado a partir de texto y/o imagen.	Creación de contenidos audiovisuales generados por IA.
	Movie Gen (Meta)	Modelo que produce vídeos completos con sonido desde textos descriptivos.	Desarrollo de vídeos personalizados con audio.
Multimodal (Texto + Imagen + Audio + Vídeo)	Gemini (Google)	Modelo multimodal que procesa y genera contenido en diferentes formatos simultáneamente.	Asistentes que combinan texto, imagen, voz y vídeo.

Copyright 2026

Modalidad / Uso	Modelo / Ejemplo	Descripción breve	Aplicación típica
	Modelos multimodales unificados (investigación)	Investigaciones académicas en modelos que integran texto, imagen, audio y vídeo.	Proyectos de IA generalistas para múltiples tareas.

Los **modelos fundacionales** son redes neuronales entrenadas con datos masivos y diseñadas para poder **adaptarse a numerosas tareas** tanto de generación como de comprensión en diferentes dominios de entrada y salida (texto, imagen, audio y vídeo). Su carácter generalista los convierte en la base sobre la que se construyen aplicaciones especializadas en IA.

Revision #9
Created 2026-02-18 17:54:08 CET by Luis Hueso
Updated 2026-03-18 16:10:42 CET by Maria

Copyright 2026