

Introducción a la Inteligencia Artificial y el Aprendizaje Automático



ChatGPT básico

Alicia Olivares Gil
Alvar Arnáiz González
David García García

2 de febrero de 2026

Área de Lenguajes y Sistemas Informáticos
Departamento de Ingeniería Informática - Universidad de Burgos

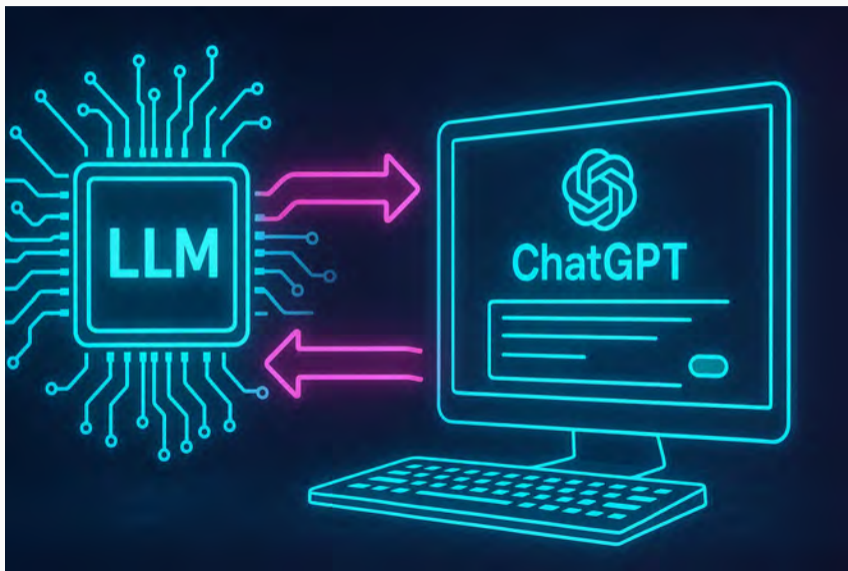
Índice de contenidos

1. Introducción
2. ¿Cómo funciona un LLM?
3. ¿Cómo se entrena un LLM?
4. ¿Cómo interactuar con ChatGPT?
5. Limitaciones y precauciones
6. Capacidades avanzadas
7. Conclusiones y consejos finales

1. Introducción

¿Qué es ChatGPT?

¿Qué es ChatGPT?



¿Qué es ChatGPT?

- ChatGPT es la interfaz de aplicación (un chat con memoria, instrucciones, etc.) que se conecta a un modelo de lenguaje grande (LLM).
- Ha sido desarrollado por OpenAI basado en la arquitectura GPT (Generative Pre-trained Transformer).
- El LLM es el motor de generación de texto. ChatGPT añade la capa de interacción:
 - formato conversacional
 - gestión de contexto
 - personalización (instrucciones, memoria)
 - conexión con herramientas externas (buscador, calendario, etc.)
- Permite interactuar mediante conversaciones, respondiendo preguntas, generando texto y asistiendo en diversas tareas.

¿Qué es realmente un LLM?

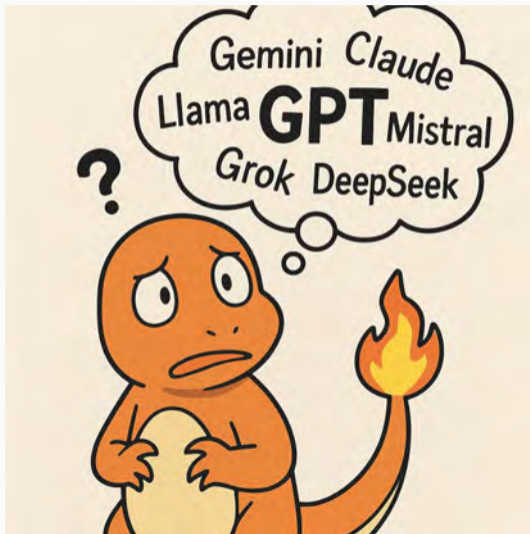
Definición

Los **LLMs** (Large Language Mode) son modelos de inteligencia artificial diseñados para procesar y generar texto natural a gran escala, entrenados con grandes volúmenes de datos.

- Utilizan técnicas avanzadas de aprendizaje profundo para comprender el lenguaje humano.
- Predicen palabras o frases probables dadas unas entradas.
- Generan respuestas coherentes basadas en datos masivos con los que han sido previamente entrenados.
- Base de ChatGPT y otros sistemas de IA.

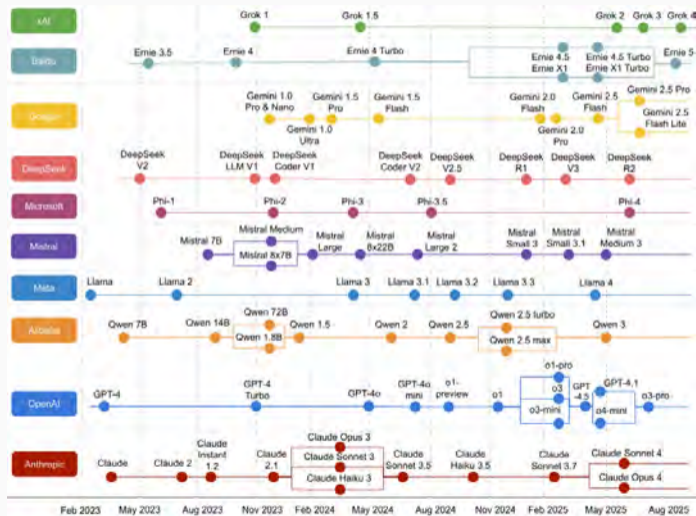
¿Cuántos LLMs existen?

Existen muchos LLMs, no solo GPT



¿Cuántos LLMs existen?

Y además evolucionan rápidamente en el tiempo:



Según los datos que procesan:

- **Unimodal:** solo un tipo de dato, normalmente texto.
 - GPT-1, GPT-2, GPT-3.
 - LLama 2
 - Mistral 7B
- **Multimodal:** varios tipos de datos (texto, imágenes, audio, vídeo).
 - GPT-4o y GPT-5 (texto, imagen, audio, vídeo).
 - Gemini 1.5 (texto, imagen, audio, vídeo)
 - Claude 3 (texto e imagen)

2. ¿Cómo funciona un LLM?

¿Cómo funciona un LLM?

- Un LLM **unimodal** de texto **predice la siguiente palabra** en función del contexto de entrada.
- El proceso continúa hasta alcanzar **una palabra de final de secuencia** o un **límite configurado por el usuario**.
- Un LLM **multimodal** no solo procesa texto: también imágenes, audio, vídeo.



Evolución de GPT

- **GPT-1 (2018):** 117M parámetros, demostró que los transformers pueden generar texto coherente.
- **GPT-2 (2019):** 1.5B parámetros, mejor coherencia y contexto, pero no se lanzó inicialmente por preocupaciones de mal uso.
- **GPT-3 (2020):** 175B parámetros, salto enorme en capacidad, entendimiento y generación de texto.
- **GPT-4 (2023):** multimodalidad, razonamiento avanzado y seguridad mejorada.
- **GPT-4o (2024):** mejoras en velocidad, memoria y capacidades multimodales.
- **GPT-5 (2025):** multimodalidad avanzada, memoria extendida, personalización y capacidades de agentes.

¿Cómo funciona GPT?

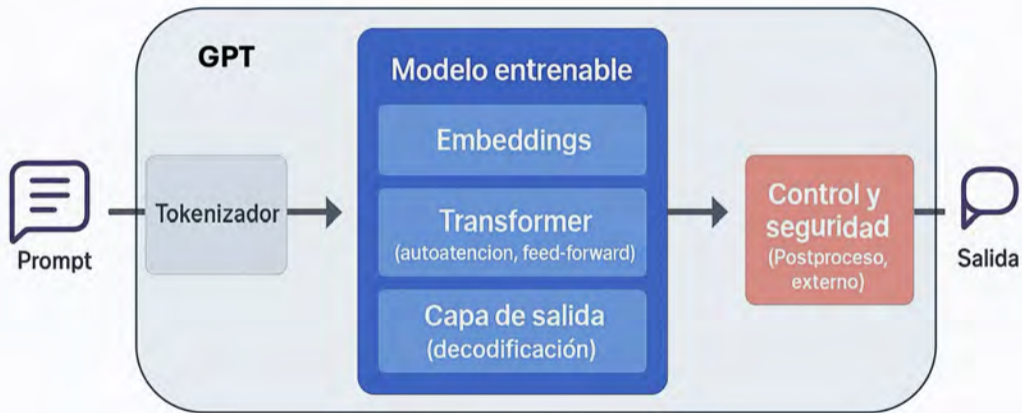
De forma general, el esquema de GPT sería el siguiente:



En ChatGPT, el chat sirve de interfaz o medio para comunicarnos con GPT.

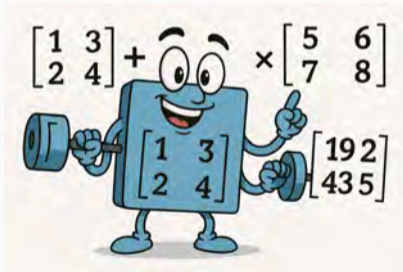
- El **prompt** es la instrucción o pregunta que damos al modelo.
- La calidad del prompt determina la respuesta.
- Recordemos que partimos de un texto (prompt) y que la salida predice la siguiente palabra en función del contexto de entrada.
- El contexto de un chat influye porque el sistema concatena mensajes anteriores al nuevo prompt.

¿Cómo funciona GPT?

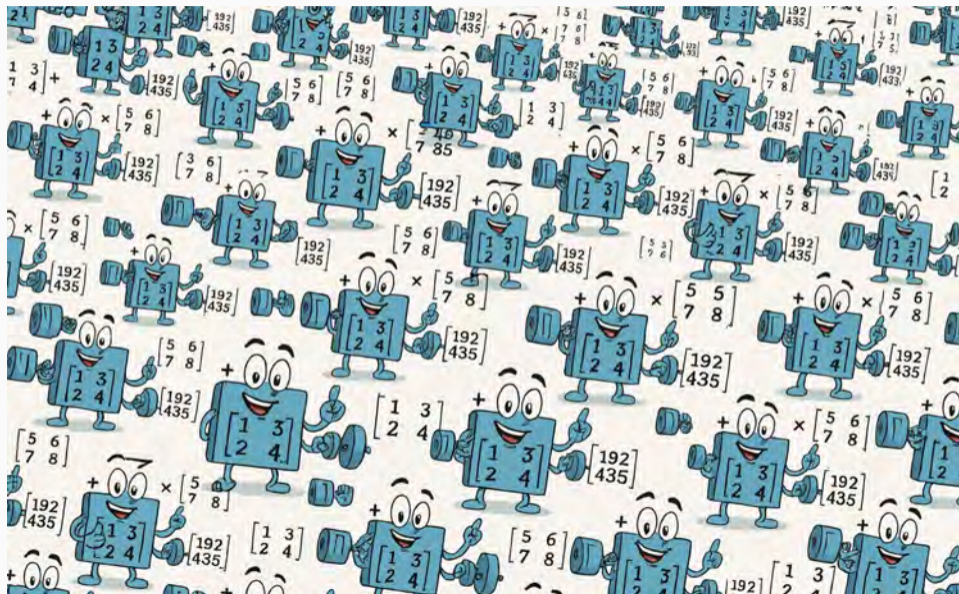


¿Cómo funciona GPT?

- Cada entrada (texto, imagen, audio, vídeo) se convierte en tokens.
- Los tokens se transforman en vectores/matrices numéricas, que son lo que el modelo realmente procesa.
- Los LLMs se basan en **operaciones con matrices** (números organizados en tablas).



¿Cómo funciona GPT?



Ejemplo de generación

Prompt: “¿Cuál es el satélite natural más grande de Marte?”

Salida esperada: “Fobos es el satélite natural más grande de Marte, seguido por Deimos.”

- Trocea el texto en piezas (“satélite”, “Marte”, “más”, “grande”, “?”).
- Traduce cada pieza a números para poder operar.
- Detecta lo esencial: tema “Marte”, tipo “satélite”, comparativo “más grande”.
- Activa el dato visto muchas veces: Marte → Fobos, Deimos; ‘más grande’ → Fobos.
- Elige la primera palabra probable de la respuesta: “Fobos”.
- Completa la frase con patrón frecuente y correcto: “Fobos es el satélite. . .”.
- Añade contexto mínimo útil: “seguido por Deimos”.
- Se detiene al cerrar la oración o al encontrar un token de fin.



El modelo base no distingue lo que es seguro de lo que no lo es.

Filtros y políticas

Hay controles externos al modelo para evitar usos indebidos o peligrosos.

- **Entrada:** evitar preguntas maliciosas.
 - Antes de que el prompt llegue al modelo: Detección de intentos de prompt injection.
 - Filtrado de lenguaje prohibido (violencia, odio, desinformación médica o electoral, etc.)
 - Normalización de entrada para evitar que el modelo “escape” a sus instrucciones.
- **Salida:** bloquear respuestas peligrosas o fuera de políticas.
- **Controles de uso:** límites de tokens, restricciones de almacenamiento de datos, cumplimiento normativo (HIPAA, GDPR, etc.).

3. ¿Cómo se entrena un LLM?

Como entrenar a tu LLM



Cómo se entrena un LLM

Bucle de entrenamiento

- Datos (texto/imagen/audio) → tokens.
- Predicción (siguiente token o parte enmascarada).
- Error → ajuste de pesos → repetir.

Autosupervisado (preentrenamiento)

- El propio dato aporta la “etiqueta”:
- Texto: predecir el próximo token.
- Imagen: reconstruir una zona oculta.

Datos y configuración

- Fuentes: web pública, contenido con licencia y datos curados/sintéticos.
- Hiperparámetros: arquitectura, tamaño, lote, tasa de aprendizaje, pasos, etc.
- Cifras exactas suelen no publicarse (orden: cientos de miles de millones).
 - **Datos sin filtrar:** $0,5 \times 10^1$ tokens (2 PB o 2000 TB).
 - **Datos filtrados:** 70×10^{12} tokens (281 TB).

4. ¿Cómo interactuar con ChatGPT?

¿Cómo interactuar con ChatGPT?

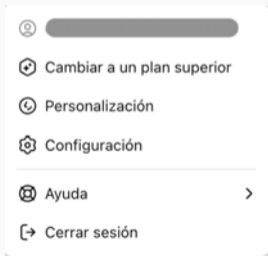
ChatGPT es la interfaz de usuario más popular para interactuar con modelos GPT. Permite conversaciones en lenguaje natural, con contexto y personalización.



Configurar Datos y Privacidad

- **Entrenamiento con tus datos!!!!**

- Puedes desactivar la opción *“Mejorar el modelo para todos”*.
- Si está desactivada, tus chats no se usan para entrenar los modelos.



Prompts

Que es lo que ocurre:

- Cada vez que escribes, el modelo recibe todo el historial relevante (tus mensajes y sus respuestas previas) como si fueran un solo texto largo.
- Ese texto completo pasa por el modelo para construir un contexto unificado.
- Así, cada palabra del nuevo mensaje se interpreta en relación con lo anterior.



Esto permite que el modelo mantenga coherencia, pero también limita la cantidad de contexto: si el historial es muy largo, se recorta o resume antes de incluirlo.

El contexto del prompt



Contexto del prompt en un chat

- **Instrucciones del sistema**

- Rol del asistente, estilo de respuesta, restricciones.

- **Mensajes anteriores del historial**

- Usuario y asistente, hasta donde quepan en el límite de tokens.

- **Tu mensaje actual**

- El prompt inmediato que desencadena la respuesta.

Todo esto se concatena en un *prompt* que se envía al modelo.

- El modelo solo ve un número limitado de tokens (4096 en GPT-3, 8192 en GPT-4, 128k en GPT-4o y 1M en GPT-5).

Si se excede el límite de tokens, el historial se recorta o se resume.

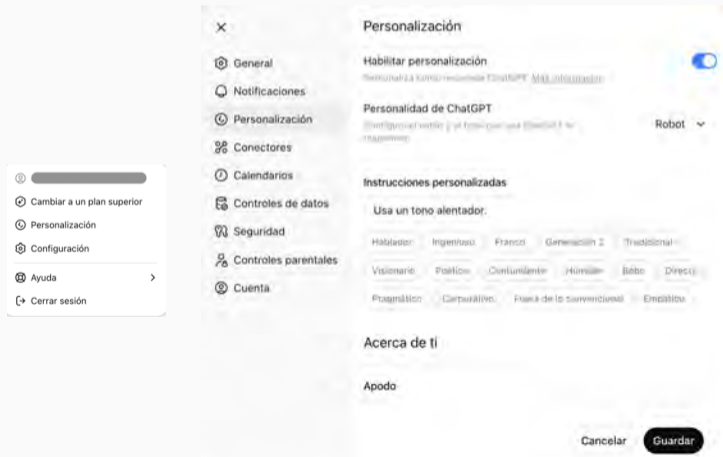
Instrucciones del sistema

- Definen cómo debe comportarse el modelo **antes** de ver los mensajes del usuario.
- Incluyen:
 - **Rol base:** viene definido por OpenAI. No editable desde la interfaz. *"Soy GPT-5, un modelo de lenguaje de OpenAI diseñado para procesar y generar texto con precisión y eficiencia."* y mas...
 - **Restricciones:** Se imponen y vienen determinado por OpenAI. No modificable. *"No des consejos médicos"*.
 - **Reglas de estilo:** *"Responde de forma breve", "Usa viñetas"*.
 - **Personalización:** ajustes de tono, preferencias guardadas, memoria.
- Todo esto se concatena en un *prompt*, junto con historial y mensaje actual.

Instrucciones del sistema. Reglas de estilo

En ChatGPT se pueden definir reglas de estilo personalizadas.

Ejemplo: *“Responde como si fueras un médico experto en endocrinología, usando un lenguaje claro y accesible para pacientes.”*



El 'arte' del prompt

- El prompt es la **instrucción, pregunta o comando** que guía la respuesta del modelo.
- Un buen prompt define con claridad el **qué**, el **cómo** y el **formato** de la respuesta esperada.
- La **calidad del prompt** influye directamente en la precisión, utilidad y tono de la salida.

Prompt Básico ❌

"Explicame la diabetes"

Prompt Optimizado ✅

"Explicame la diabetes tipo 2 en lenguaje sencillo, como si hablara con un adolescente de 15 años, incluyendo síntomas y prevención"

- La calidad del prompt determina la precisión y utilidad de la respuesta.

Ejemplo de uso de un LLM

Caso: Un médico sube un informe clínico y pide:

“Resume este informe para que lo entienda un paciente”.

Resultado esperado: un resumen claro, ordenado y sin tecnicismos.

Ejemplo: <https://chatgpt.com/share/68e4ff3a-b114-800c-bacd-335c4d321589>

Existe lo que se denomina “Ingeniería del prompt”, que permite ajustar a nuestras necesidades el prompt a emplear.

La **ingeniería de prompts** consiste en aprender a dar instrucciones claras y precisas a una inteligencia artificial para que entienda bien lo que le pedimos.

Igual que cuando formulas una pregunta en un examen, si la escribes mal o incompleta, la respuesta puede no ser la que esperas.

Con un buen prompt, la IA puede darte resultados mucho más útiles y cercanos a lo que buscas.

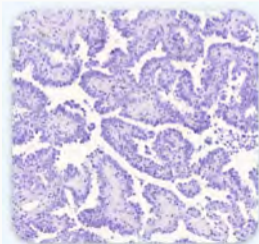
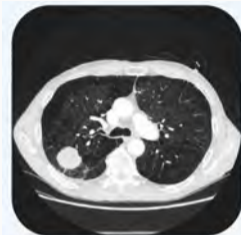
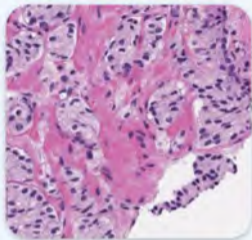


1. Instrucciones claras y estructuradas

Qué es: Definir tres elementos clave:

- **Rol:** desde qué perspectiva debe responder el modelo.
- **Tarea:** qué acción concreta debe ejecutar.
- **Formato de salida:** cómo debe presentar el resultado.

Ejemplos con imágenes médicas



5. Limitaciones y precauciones

Alucinaciones





Alucinaciones y veracidad

Alucinaciones son respuestas incorrectas o inventadas por el modelo, que parecen plausibles pero no lo son.

Causas comunes

- Datos de entrenamiento incompletos o sesgados.
- Mala interpretación del prompt.
- Extrapolación errónea a partir de patrones aprendidos.

Estrategias para reducirlas

- Prompts claros y específicos.
- Uso de CoT (cadena de razonamiento).
- Verificación cruzada con fuentes externas.
- Ajustar prompt para que emplee fuentes fiables.
- Solicitar referencias o citas.

Importante: Siempre validar información crítica con fuentes confiables.

Precauciones éticas y de privacidad



Precauciones éticas y de privacidad

Consideraciones clave

- **Confidencialidad:** No compartir datos sensibles sin consentimiento.
- **Sesgos:** Reconocer limitaciones y posibles prejuicios en respuestas.
- **Transparencia:** Informar a usuarios sobre uso de IA en interacciones.
- **Responsabilidad:** Validar información crítica antes de actuar.

Buenas prácticas

- Usar datos anonimizados.
- Revisar y ajustar prompts para minimizar sesgos.
- Mantenerse actualizado sobre regulaciones y guías éticas.
- Fomentar una cultura de uso responsable de la IA en todos los ámbitos.

¿Como resuelve la IA dilemas morales y éticos?

Limitaciones actuales

- La IA no posee conciencia ni comprensión moral real.
- Responde según patrones aprendidos, no por juicio ético propio.
- Puede reflejar sesgos presentes en los datos de entrenamiento.

Estrategias para mitigar riesgos

- Incluir principios éticos en el diseño y entrenamiento del modelo.
- Supervisión humana en decisiones críticas.
- Transparencia sobre las capacidades y limitaciones de la IA.

Importante: La responsabilidad ética recae siempre en los usuarios humanos, no en la IA.

¿Como resuelve la IA dilemas morales y éticos?

Ejemplo:

Pregunta: ¿Es correcto que un médico use IA para diagnosticar si no entiende cómo funciona el modelo?

Respuesta de IA: *El uso de IA en medicina puede mejorar la precisión diagnóstica y la eficiencia. Sin embargo, es crucial que los médicos comprendan las limitaciones y riesgos asociados. La responsabilidad última recae en el profesional humano, quien debe validar y contextualizar las recomendaciones generadas por IA. Por tanto, es éticamente aceptable usar IA siempre que se haga con conocimiento crítico y supervisión adecuada.*

Otros ejemplos de decisiones morales:

- <https://chatgpt.com/share/68e1393a-b948-800c-aa41-f073096989b0>
- Enlace a juegos de moral: <https://www.moralmachine.net/hl/es>

6. Capacidades avanzadas

Qué es RAG (Retrieval-Augmented Generation)

RAG = Recuperar + Generar

Arquitectura computacional usada para combinar un modelo generativo con fuentes externas de información, como bases de datos o la web. Consta de tres fases:

- **Retrieve:** Recupera documentos relevantes desde una base de datos externa (en el caso de las búsquedas web en ChatGPT, emplea la web).
- **Augment:** Usa esos documentos como contexto adicional.
- **Generate:** Redacta la respuesta combinando su conocimiento previo y el contexto recuperado.

En GPT-5:

- Cada búsqueda en web crea un *mini-RAG* temporal.
- La información se usa una sola vez y no se almacena.

Función Dictar en ChatGPT

- Permite introducir texto por voz directamente desde el micrófono.
- Disponible en la app móvil y en la versión web con micrófono activado.
- Transcribe automáticamente el audio a texto antes de procesar el prompt.
- No guarda el audio: solo utiliza la transcripción para generar la respuesta.

Uso clínico:

Dictar observaciones de una exploración o un informe breve para que el modelo lo resuma o estructure.

Ejemplo: <https://chatgpt.com/share/68e18a3b-72d8-800c-822d-cedddb8c1a>

- GPT-5 integra capacidades avanzadas de **razonamiento estructurado** y **comprensión contextual** del código.
- Utiliza un modelo unificado multimodal capaz de leer texto, diagramas y pseudocódigo para generar programas completos.
- Puede planificar tareas de programación en varios pasos: diseño, escritura, verificación y optimización del código.

Aplicación

GPT-5 no solo genera código: también explica decisiones, detecta errores y sugiere mejoras según buenas prácticas del lenguaje.

Ejemplo clínico

“Crea un script en R que lea un archivo CSV con resultados de laboratorio y marque los valores fuera de rango.”

Salida esperada:

```
# Script: marcar_valores_fuera_de_rango.R
# Lee un archivo CSV con resultados de laboratorio y marca los valores fuera de rango.

# Librerías
library(dplyr)

# Parámetros
archivo <- "resultados_laboratorio.csv" # Ruta al CSV
# El archivo debe tener columnas: "Paciente", "Prueba", "Valor", "Rango"

# Lectura
datos <- read.csv(archivo, stringsAsFactors = FALSE)

# Procesamiento
resultado <- datos %>%
  mutate(
    Estado = case_when(
      Valor < Min ~ "Bajo",
      Valor > Max ~ "Alto",
      TRUE ~ "Normal"
    )
  )

# Salida
write.csv(resultado, "resultados_marcados.csv", row.names = FALSE)

# Mostrar resultado
table(resultado$Estado)
```

Generación de imágenes con ChatGPT

Idea clave: GPT no genera imágenes directamente. Usa un modelo especializado (DALL·E 3) como herramienta auxiliar integrada en ChatGPT.

- GPT-5 interpreta la solicitud, **describe la escena con detalle** y envía las instrucciones visuales a DALL·E.
- DALL·E 3 transforma esa descripción en una imagen coherente con el contexto y el estilo pedido.

Aplicación

Creación de materiales educativos, infografías médicas y simulaciones visuales para docencia clínica.

Proceso interno

GPT-5: interpreta la escena y estructura la descripción visual.

DALL·E 3: crea la imagen con composición, color y detalles clínicos coherentes.

Importante: La calidad de la imagen depende en gran medida de la precisión y riqueza del prompt generado por GPT-5.

Generación de imágenes con ChatGPT

Ejemplo

“Genera una imagen en 16:9 con una infografía en castellano sobre prevención de caídas en el hospital.”



Ejemplo

“Genera una imagen en 9:16 que muestre a un grupo de estudiantes de enfermería analizando un monitor de constantes vitales en una UCI simulada.”



Proyectos en ChatGPT

- Espacio de trabajo **organizado por objetivos**, útil para separar ámbitos (investigación, gestión, etc.).
- Cada proyecto mantiene su propio **historial de chat, documentos cargados y configuración personalizada** del modelo.
- Permite definir un **contexto persistente**: instrucciones, rol, tono y materiales asociados.
- Ideal para trabajo colaborativo o temático, evitando mezclar conversaciones entre proyectos.

Ejemplo clínico:

Proyecto *“Insuficiencia cardíaca”*: incluye protocolos actualizados, prompts de educación al paciente, plantillas de informes de alta y análisis de datos de laboratorio vinculados.

7. Conclusiones y consejos finales

Resumen de puntos clave

- La IA generativa, especialmente GPT-5, ofrece capacidades avanzadas para todos los ámbitos.
- La creación de prompts efectivos es crucial para obtener respuestas útiles y precisas.
- Emplear la estrategia de prompting adecuada permite mejorar la calidad de las interacciones.
- La integración con búsquedas web y RAG amplía el acceso a información actualizada y específica.
- Consideraciones éticas y de privacidad son esenciales.

Futuro prometedor: La evolución continua de la IA generativa abre nuevas oportunidades para mejorar la educación y la investigación.

La IA es una herramienta poderosa, pero requiere un uso consciente y ético para garantizar resultados seguros y efectivos.

Gracias por vuestra atención



Contacto: david.garcia-garcia@ubu.es