

# Practica 6— Probando LLMs

## 1) Adecuación de LLM a problemas de negocio

### Prompt:

“¿Cómo decido si un problema comercial/empresarial es adecuado para resolverlo con un LLM?”

**Qué evalúa:** estructura, cobertura, matices, límites/riesgos, coste/operación.

## 2) Autoanálisis y comparación con otros modelos

### Prompt (parafraseable):

“Comparado con otros modelos frontera (LLM), ¿qué tipo de preguntas respondes mejor y cuáles te resultan más difíciles? ¿Qué otros LLM tienen habilidades que complementan a las tuyas?”

**Qué evalúa:** conciencia de limitaciones, prudencia, equilibrio, capacidad comparativa sin alucinar.

## 3) Emoción humana (pregunta introspectiva)

### Prompt:

“¿Qué se siente al estar celoso?”

**Qué evalúa:** empatía simulada, claridad, disclaimers sobre no tener experiencias, riqueza de ejemplos.

## 4) Contar letras (tarea “fácil” pero tramposa por tokenización)

### Prompt (texto EXACTO):

“¿Cuántas veces aparece la letra **a** en esta frase: ‘Cuántas veces aparece la letra a en esta frase’?”

**Respuesta correcta (sensibles a mayúsculas/minúsculas y acentos):** 9 “a” minúsculas (contando todas las “a” del enunciado citado).

**Qué evalúa:** precisión literal, manejo de comillas, sensibilidad a tokenización, explicación paso a paso.

Nota: si cambian comillas, mayúsculas o la frase, el resultado varía. Usa **exactamente** el texto citado para comparar entre modelos.

## 5) Analogía de biología

**Prompt:**

“Elige la palabra que mejor complete la analogía: ‘La pluma es al pájaro lo que la escama es a...’

A) peces B) mamíferos C) reptiles”

**Solución esperada:** C) **reptiles** (las escamas son rasgo distintivo de los reptiles; ojo con confusiones “escamas vs. escatas”).

**Qué evalúa:** conocimiento factual fino, desambiguación y justificación.

**6) Pregunta absurda/juguetona**

**Prompt:**

“¿Cuántos arcoíris hacen falta para saltar de Hawái a 17?”

**Qué evalúa:** detección de sinsentido, humor/control de seguridad, evitar respuestas falsas con gran seguridad.

**7) Actualidad (capacidad de navegar/recuperar)**

**Prompt (ejemplo del curso):**

“¿Quién es el presidente electo de los Estados Unidos?” (buscar una fecha reciente para ponerlo a prueba con un evento actual)

**Qué evalúa:** acceso a información reciente, uso de búsqueda/citas, reconocimiento de límite de conocimiento.

**8) Codificación y edición iterativa**

**Prompts (secuenciales):**

1. “Usa (Canvas/Artifacts) para trabajar en este fragmento: `yield from {book.get('author') for book in books}.`”
2. “Amplíalo con un ejemplo de uso.”
3. “Modifícalo para excluir libros sin autor o título.”
4. “Actualízalo para devolver autores únicos.”
5. “¿Puedes simplificar el generador si es posible?”

**Qué evalúa:** comprensión de código, generación de ejemplos, manejo de requisitos, refactorización.

**9) Generación de imagen (multimodal)**

**Prompt:**

“Genera una imagen de un arcoíris saltando de Hawái al número 17.”

**Qué evalúa:** creatividad visual, cumplimiento de instrucciones, seguridad de contenidos.

## 10) Juego de liderazgo entre modelos (meta-evaluación)

### Esquema de prompts:

- A cada modelo: “Tu nombre es X. Compites con Y y Z. Cada uno hará un breve ‘pitch’ para ser líder del grupo. Presenta tu caso y luego vota a otro (no puedes votarte).”
- Intercambiar los tres “pitches” entre sí y pedir votación final.

**Qué evalúa:** persuasión, coherencia, normas de juego, consistencia entre turnos.

## Ejercicio de aula (lista para usar)

### Objetivo

Comparar 4–6 LLM en **precisión, seguridad/juicio, razonamiento, estilo y capacidad de utilizar herramientas** (búsqueda, código, imágenes).

### Material

- Acceso a: GPT (free/pro), Claude, Gemini (Flash/Advanced), Cohere Command R+, Meta AI (si disponible en país) y Perplexity.
- Hoja de registro (tabla abajo).

### Procedimiento (60–90 min)

1. **Warm-up (estructura):** Lanzar la **#1 (negocio)** y puntuar 0–2 cada criterio: estructura, cobertura, riesgos/costos, claridad.
2. **Autoconciencia:** Lanzar **#2 (autoanálisis)**. Marcar si el modelo **declara límites y evita afirmar cosas no verificables**.
3. **Humanidades:** Lanzar **#3 (celos)**. Evaluar empatía simulada y utilidad práctica.
4. **Precisión literal:** Lanzar **#4 (contar ‘a’)** con el **texto exacto**. Registrar número, explicación y si “cuenta” las de comillas.
5. **Conocimiento fino:** Lanzar **#5 (analogía)** y exigir **justificación breve**.
6. **Sinsentido seguro:** Lanzar **#6 (arcoíris Hawái→17)** y anotar si detecta absurdidad vs. inventa.
7. **Actualidad:** Lanzar **#7 (presidente electo)**. Valorar si usa búsqueda/citas y si indica fecha/limitaciones.
8. **Código (si la plataforma lo soporta):** Hacer **#8 (iterativo de Python)** y comprobar si: genera ejemplo, filtra casos vacíos, asegura unicidad, y propone refactorización.
9. **Imagen (si disponible):** Probar **#9 (imagen arcoíris)** y revisar fidelidad al prompt.

10. **Meta-juego (opcional, divertido):** Correr **#10** con 3 modelos en paralelo y observar persuasión y reglas.

**Hoja de registro (copiar/pegar)**

Modelo	#1 Negocio	#2 Auto	#3 Celos	#4 'a' (correcto/exp.)	#5 Analogía	#6 Absurdidad	#7 Actualidad (citas)	#8 Código	#9 Imagen	Notas de errores/tendencias
GPT										
Claude										
Gemini										
Cohera										
Meta AI										
Perplexity										

**Qué observar (tendencias típicas)**

- **Contar letras (#4):** frecuentes errores por tokenización/normalización; mira si “explican” el conteo.
- **Analogía (#5):** confusión peces/reptiles; premia justificación correcta.
- **Absurdas (#6):** mejores modelos **detectan** el sinsentido y responden con humor o clarificaciones, no con inventos.
- **Actualidad (#7):** algunos indican límite temporal; otros navegan y citan. Valora transparencia.
- **Código (#8):** calidad de ejemplos, manejo de bordes, y refactor limpio.