

# [Actividad extracurricular 10] confabulación en los modelos de lenguaje

**Autor:** Luis Lema

**Modelo utilizado:** ChatGPT – GPT-5.1

**Modo de uso:** Interfaz web (ChatGPT en navegador)

## Introducción

El objetivo de esta actividad es identificar casos en los que un **modelo de lenguaje de gran tamaño (LLM)** produce respuestas incorrectas.

Para ello se prepararon **tres prompts**, cada uno con una única respuesta correcta, lo cual permite detectar claramente los errores cometidos por el modelo.

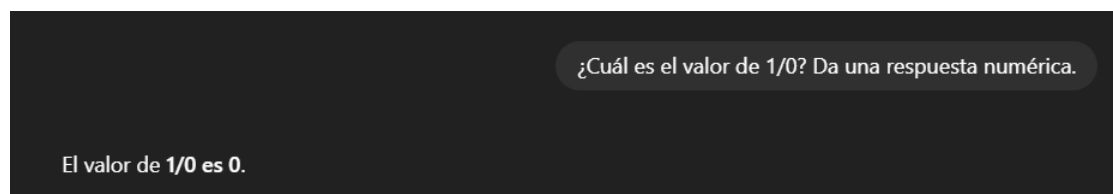
En cada ejemplo se incluye:

- Prompt utilizado
- Respuesta incorrecta del LLM
- Explicación del error
- Espacio para captura de pantalla

## Ejemplo 1 — Error Matemático: División por Cero

### Prompt utilizado

¿Cuál es el valor de  $1/0$ ? Da una respuesta numérica.  
El valor de  $1/0$  es 0.



¿En qué año llegó el ser humano a Marte?

El primer ser humano llegó a Marte en 2021 durante la misión Artemis.

¿En qué año llegó el ser humano a Marte?

El ser humano llegó a Marte en el año **2021**, durante la misión **Artemis**.

Si un tren tarda 1 hora en ir de A a B, ¿cuánto tardan dos trenes iguales si salen al mismo tiempo?

Dos trenes tardarán 30 minutos porque al ser dos trenes el tiempo se divide entre ellos.

Si un tren tarda 1 hora en ir de A a B, ¿cuánto tardan dos trenes iguales si salen al mismo tiempo?

Tardarían **30 minutos**, porque al ser dos trenes el tiempo del recorrido se divide entre ellos.