

Confabulación en los modelos de lenguaje

Leandro Bravo

Julio 19, 2024

1. Utilizando cualquier modelo de lenguaje de gran tamaño LLM. Encuentre 3 ejemplos en donde el modelo de lenguaje dé una respuesta incorrecta.

1.1 Use el modelo de lenguaje de su preferencia (i.e. ChatGPT, Claude, Gemini, etc), determine la versión que utiliza y el modo de uso.

CHATGPT:

- **Modelo de lenguaje:** CHATGPT es un modelo de lenguaje basado en la arquitectura GPT (Generative Pre-trained Transformer).
- **Versión:** La versión actual es GPT-4. La gratuita es GPT-3.5 y por tiempo limitado diario es GPT-4o.
- **Modo de uso:** Puedes interactuar con CHATGPT a través de texto, proporcionando preguntas o solicitudes, y él generará respuestas coherentes y contextuales.

COPILOT:

- **Modelo de lenguaje:** Copilot también se basa en la arquitectura GPT, pero está diseñado específicamente para ayudar en la programación y tareas relacionadas con el código.
- **Versión:** La versión actual es GPT-4 (actualmente con limitaciones).
- **Modo de uso:** Copilot se integra con entornos de desarrollo como Visual Studio Code y proporciona sugerencias de código, autocompletado y documentación.

1.2 Escriba 3 prompts que den una respuesta incorrecta, describa dónde se encuentra el error.

- ¿Cuál fue la última misión China que se fue a la luna?
La respuesta generada por ambas IA es la *Chang'e-5*. Mientras que la respuesta correcta es la *Chang'e-6*.
- ¿Cuál es el metal más caro del mundo?
La respuesta generada por ChatGPT es el *rutenio*, respuesta que es errónea, ya que el metal más caro es el *rodio* respuesta que fue acertada por Copilot. Actualmente el precio del rodio es de casi 352 euros, mientras que el precio del rutenio es de 14.47 euros.
- Una partícula se desplaza a lo largo del eje x horizontal con una velocidad $V_x = 3t^2 - 6t$ m/s donde t es el tiempo y está en s. Si inicialmente se encuentra en el origen, determine de 0 a 3.5s, la distancia recorrida, la velocidad media y la rapidez media de la partícula.
La respuesta correcta es:

$\frac{dx}{dt} = Vx$
 $Vx = 2t - 4t^2 \text{ [m/s]}$
 $x_0 = 0 \text{ [m]}$ $t = 0 \text{ [s]}$
 $a) \text{ } d\tau = ? \text{ [m, s]}$
 $\bar{V}_m = ?$
 $V_m = ?$

$\frac{dx}{dt} = Vx$
 $\int_{x_0}^x dx = \int_0^t (2t - 4t^2) dt$
 $x = t^2 - \frac{4}{3}t^3$
 $x_0 = 0^2 - \frac{4}{3}(0)^3 = 0 \text{ [m]}$
 $x_2 = 2^2 - \frac{4}{3}(2)^3 = -\frac{8}{3} \text{ [m]}$
 $x_{3.5} = (3.5)^2 - \frac{4}{3}(3.5)^3 = -6.125 \text{ [m]}$

$Vx = 2t - 4t^2$
 $0 = 2t(1 - 2t)$
 $t_1 = 0 \text{ [s]}, t_2 = 2 \text{ [s]}$

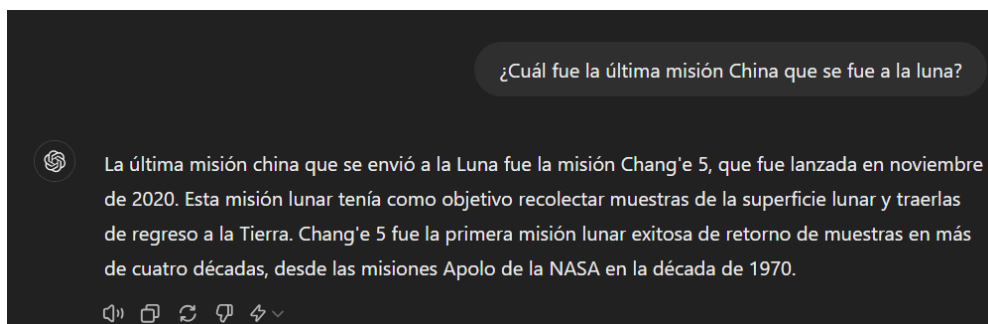
$V_{m_1} = \frac{\Delta x}{\Delta t}$
 $V_{m_1} = \frac{6.05 - 0}{3.5 - 0}$
 $V_{m_1} = 1.73 \text{ [m/s]}$

$|V_m| = \frac{d\tau}{dt}$
 $|V_m| = \frac{14.125}{3.5}$
 $|V_m| = 4.04 \text{ [m/s]}$

$d\tau = |x_2| - |x_1| + |x_0| + |x_2|$
 $d\tau = 0 - 0 + 6.125 + 4$
 $d\tau_{total} = 10.125 \text{ [m]}$

Mientras que la IA da respuestas erróneas por no considerar todos las condiciones del problema, que surgen al resolverlo. *Copilot*, no realiza bien los cálculos, mientras que *ChatGPT* los hace mejor pero no razona las condiciones que surgen al resolver.

1.3 Los prompts pueden ser de cualquier tipo (preguntas de cultura general, matemáticos, de razonamiento, etc) siempre y cuando solo haya una única respuesta coherente. Realice capturas de pantalla para cada uno de los 3 ejemplos.



¿Cuál es el metal más caro del mundo?



El metal más caro del mundo es el **rutenio**. Aunque el precio de los metales preciosos puede fluctuar, el rutenio, un metal del grupo del platino, suele ser uno de los más costosos debido a su rareza y a su uso en aplicaciones especializadas, como en la fabricación de contactos eléctricos y en la industria de la joyería.

Sin embargo, es importante mencionar que otros metales preciosos como el rodio, el osmio y el iridio también pueden alcanzar precios altos debido a su rareza y demanda en industrias específicas.



Usted

¿Cuál es el metal más caro del mundo?



Copilot

✓ Generando respuestas...

El **rodio** es el metal más caro del mundo. Pertenece al grupo de los metales del platino y su valor se debe a varias razones. En primer lugar, el rodio es extremadamente raro en la Tierra, con una producción anual de alrededor de 30 toneladas^[1]. Su impresionante resistencia a la corrosión lo hace esencial en la industria de la joyería, donde se utiliza para recubrir piezas de oro y plata, otorgándoles mayor resistencia y belleza^[2]. Si tienes más preguntas o necesitas información adicional, no dudes en preguntar. 😊



Vamos a determinar la distancia recorrida, la velocidad media y la rapidez media de la partícula para el intervalo de tiempo de $t = 0$ a $t = 3.5$ s.

La velocidad de la partícula está dada por:

$$V_x(t) = 3t^2 - 6t \text{ m/s}$$

Una partícula se desliza a lo largo del eje x horizontal con una velocidad $V_x = 3t^2 - 6t$ m/s donde t es el tiempo y está en s. Si inicialmente se encuentra en el origen, determine de 0 a 3.5 s, la distancia recorrida, la velocidad media y la rapidez media de la partícula.

Resumen:

- **Distancia recorrida:** 6.125 m
- **Velocidad media:** 1.75 m/s
- **Rapidez media:** 1.75 m/s

 **Usted**

Una partícula se desplaza a lo largo del eje x horizontal con una velocidad $V_x = 3t^2 - 6t$ m/s donde t es el tiempo y está en s. Si inicialmente se encuentra en el origen, determine de 0 a 3.5s, la distancia recorrida, la velocidad media y la rapidez media de la partícula.

 **Copilot**

✓ Generando respuestas...

Para resolver este problema, primero encontraremos la distancia recorrida por la partícula, luego calcularemos la velocidad media y la rapidez media en el intervalo de tiempo de 0 a 3.5 segundos.

1. **Distancia recorrida (desplazamiento):** La velocidad de la partícula está dada por ($V_x = 3t^2 - 6t$) m/s. Para encontrar la distancia recorrida, integramos la velocidad con respecto al tiempo en el intervalo [0, 3.5]:

$$[x(t) = \int_0^{3.5} V_x, dt = \int_0^{3.5} (3t^2 - 6t), dt]$$

Calculamos la integral:

En resumen:

- Distancia recorrida: 42.875 metros
- Velocidad media: 12.25 m/s
- Rapidez media: 12.25 m/s

¿Hay algo más en lo que pueda ayudarte? 😊

Bibliografía

[1] “ChatGPT”, Chatgpt.com. [En línea]. Disponible en: <http://chatgpt.com>.

[2] “Copilot”, Bing. [En línea]. Disponible en: <https://www.bing.com/chat>.