

EXPLORACIÓN DEL POTENCIAL DIDÁCTICO DE LAS ALUCINACIONES DE CHATGPT

Andrés Merino

Junio 2025

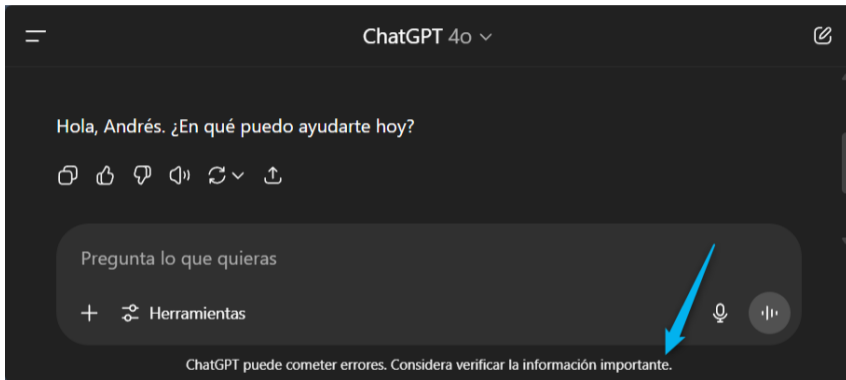
Facultad de Ciencias Exactas, Naturales y Ambientales



CONTENDIO

1. Introducción
2. ¿Cómo funciona ChatGPT?

INTRODUCCIÓN



ChatGPT puede cometer errores: ¿prohibimos su uso en aula o lo aprovechamos para enseñar?

Objetivo de la charla

- Reflexionar sobre el funcionamiento de ChatGPT y su tendencia a generar respuestas incorrectas.
- Presentar una experiencia concreta en el aula donde las alucinaciones se usaron como recurso didáctico.
- Explorar el diseño de GPTs personalizados que inducen errores con fines pedagógicos.

Objetivo de la charla

- Reflexionar sobre el funcionamiento de ChatGPT y su tendencia a generar respuestas incorrectas.
- Presentar una experiencia concreta en el aula donde las alucinaciones se usaron como recurso didáctico.
- Explorar el diseño de GPTs personalizados que inducen errores con fines pedagógicos.

Caso de uso: Cálculo Diferencial

¿CÓMO FUNCIONA CHATGPT?

¿CUÁL ES LA PRÓXIMA PALABRA?

El niño fue al

→ **Café**

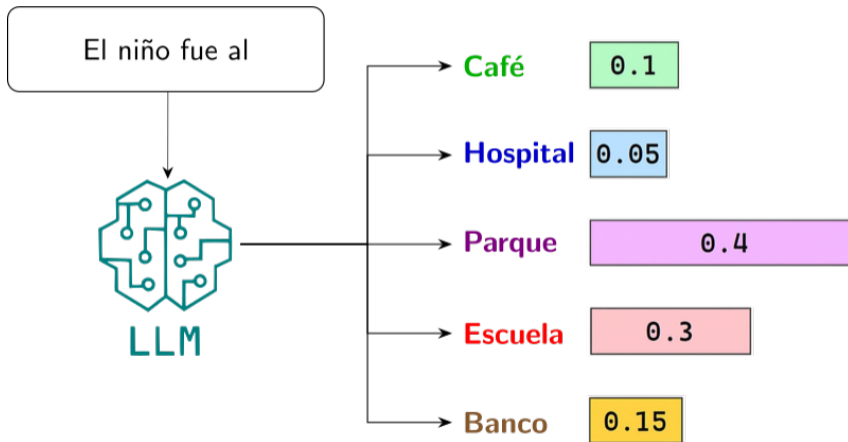
→ **Hospital**

→ **Parque**

→ **Escuela**

→ **Banco**

¿CUÁL ES LA PRÓXIMA PALABRA?

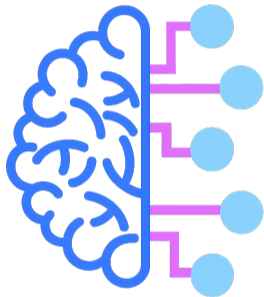


¿CÓMO FUNCIONA CHATGPT?

- ChatGPT es una herramienta que usa los Modelos GPT de OpenAI.
- Los modelos GPT de OpenAI son **modelos grandes de lenguaje** (*Large Language Model*, LLM).
- Su tarea principal es predecir la **palabra más probable** dada una secuencia anterior.
- No comprende el significado, solo calcula probabilidades a partir de patrones lingüísticos aprendidos.



PREDICCIÓN DE LA SIGUIENTE PALABRA



- El modelo asigna una **distribución de probabilidades** a cada posible palabra siguiente.
- No elige al azar: selecciona las opciones con mayor probabilidad.
- Por eso puede generar respuestas coherentes... o también **erróneas** de manera convincente.

¿CÓMO SE ENTRENÓ CHATGPT?

- GPT-3 fue entrenado con más de **300 mil millones de tokens** (~570 GB de texto limpio).
- Para GPT-4 se utilizó un volumen mucho mayor, aunque OpenAI no ha revelado cifras exactas.

Fuentes principales

- **Common Crawl** (filtrado y depurado).
- **WebText2, Books1, Books2, y Wikipedia en inglés.**

Gracias



Presentación

Contacto: aemerinot@puce.edu.ec