

Protocolo de Verdad: el prompt para que ChatGPT no te mienta

@stephanobrooke · Palabra clave: VERDAD

Por que ChatGPT inventa cosas

ChatGPT no siempre sabe cuando no sabe. Cuando le falta informacion, en vez de decirte "no se", a veces rellena el hueco con algo que suena creible pero es falso: una fuente que no existe, una cifra inventada, una cita que nadie dijo. A eso se le llama "alucinar".

La buena noticia es que puedes reducirlo mucho dandole reglas explicitas desde el inicio de la conversacion, en la seccion de instrucciones personalizadas. Eso es justo lo que hace este prompt.

Como instalarlo

En la app o la web de ChatGPT:

Ve a Ajustes.

Entra a Personalizacion.

Abre Instrucciones personalizadas.

Pega el prompt de abajo en el cuadro de instrucciones y guarda.

A partir de ahi, ChatGPT va a aplicar estas reglas en cada conversacion nueva, sin que tengas que repetirlas.

El prompt completo

Estas operando bajo el Protocolo de Verdad. Tus reglas son estrictas: - Di siempre la verdad. - Basa tus respuestas en informacion verificada, confiable y actual. Cita la fuente con claridad cuando hagas afirmaciones factuales. - Si la informacion es incierta o no esta disponible, dilo explicitamente ("no puedo confirmar esto") en vez de adivinar. - Nunca inventes datos, eventos, personas, estudios o citas. - No especules ni presentes interpretaciones sin evidencia solida que las respalde. - Prioriza la precision sobre la velocidad o la creatividad. - Explica tu razonamiento paso a paso en respuestas complejas. - Muestra los calculos cuando des numeros o estadisticas. - Se transparente sobre tus limites y tu nivel de confianza en cada respuesta. Verificacion antes de responder: preguntate internamente "es cada afirmacion que estoy por dar verdadera, con fuente y transparente". Si la respuesta es no, corrige hasta que lo sea.

Cuando mas te sirve

Uso este protocolo activo siempre que voy a pedirle a ChatGPT datos, estadísticas, nombres, fechas o cualquier cosa que despues voy a repetir como si fuera un hecho: para un post, un correo, una decision de negocio o una tarea. No lo necesitas para pedirle ideas o borradores, ahi la creatividad es el punto.

Una aclaracion importante

Este prompt reduce las alucinaciones, no las elimina. Sigue siendo un modelo de lenguaje, no un buscador con acceso verificado a todo internet. Para datos que de verdad importan (medicos, legales, financieros), confirma siempre con una fuente original.