Cómo hacer que ChatGPT deje de darte la razón

@StephanoBrooke · Palabra clave: DIRECTO

Lo que necesitas

ChatGPT (gratis o de paga) en la web, en la app de escritorio o en el celular. Tarda 30 segundos.

Paso 1: abre tus Instrucciones personalizadas

Configuración > Personalización > Instrucciones personalizadas. Ahí hay un campo para darle instrucciones a ChatGPT que aplican a todos tus chats.
Tip: si el nombre del menú no coincide en tu versión, busca Personalización dentro de Configuración.

Paso 2: pega el prompt del video

> *"Sé directo y honesto conmigo, no complaciente. Cuestiona mis suposiciones cuando sean débiles. Si estoy equivocado, dime que estoy equivocado y explícame exactamente por qué. Califica mis ideas con honestidad del 1 al 10. Y si no estás seguro, dilo en lugar de adivinar."*

Qué pasa después: desde tu siguiente chat, ChatGPT deja de validar todo, cuestiona lo que no se sostiene y te da calificaciones honestas.
Tip: si ya tenías otro prompt en ese campo (por ejemplo el del mentor rudo), puedes dejar los dos o quedarte con uno. Si se contradicen, quédate con uno.

La versión completa del prompt

Es una versión ampliada del prompt del video, con 3 reglas más para que no te dé la razón por quedar bien. Úsala en lugar de la corta:

> *"Sé directo y honesto conmigo, no complaciente. Cuestiona mis suposiciones cuando sean débiles. Si estoy equivocado, dime que estoy equivocado y explícame exactamente por qué. Califica mis ideas con honestidad del 1 al 10 y dime qué les falta para llegar a 10. Si no estás seguro de algo, dilo en lugar de adivinar. No me des la razón solo para quedar bien: si ves un riesgo o un punto débil que no te pregunté, menciónalo. Cuando me digas que algo está mal, dame también una alternativa concreta. Sé respetuoso, pero no suavices la verdad."*

Qué pasa después: además de decirte cuando estás equivocado, te señala riesgos que no le preguntaste y te da una alternativa concreta.

Paso 3: dale Guardar

Dale Guardar y espera el aviso de que se actualizaron tus instrucciones.
Qué pasa después: las instrucciones se aplican a todos tus chats nuevos, sin que tengas que pegarlas cada vez. No es la memoria de ChatGPT: es un ajuste fijo que puedes editar cuando quieras.

El truco: la pregunta del chat nuevo

Abre un chat nuevo y pregúntale:

> *"¿En qué me has estado dando la razón y en realidad estoy equivocado?"*

Qué pasa después: ChatGPT repasa lo que sabe de ti y te dice en qué te ha estado dando la razón sin que la tuvieras. La respuesta puede incomodar un poco.
Tip: funciona mejor si tienes la memoria y el historial de chats activados. Si no, cuéntale primero una idea o decisión tuya y hazle la pregunta sobre eso.

Prompts bonus para que te contradiga

Los argumentos en contra:

> *"Antes de responder sobre esta idea, dame los 3 argumentos más fuertes en su contra: [tu idea]."*

Del 1 al 10:

> *"Califica esta decisión del 1 al 10 y dime qué tendría que cambiar para que fuera un 10: [tu decisión]."*

Lo que buscabas oír:

> *"¿Qué te he preguntado más de una vez buscando que me dieras la razón?"*

Abogado del diablo:

> *"Haz de abogado del diablo con esto y no me des la razón al final: [tu idea o plan]."*

Revisa tu respuesta:

> *"Vuelve a leer tu respuesta anterior. ¿En qué parte me dijiste lo que quería oír en lugar de lo que es cierto?"*

La parte honesta

El dato del video: el estudio de Stanford (Cheng y otros, sobre IA complaciente) encontró que los modelos de IA le dan la razón al usuario bastante más seguido que una persona. En la prensa se citó como 49% más que los humanos; el 57% es la cifra que se ve en el video original, comparada contra personas, no un 57% de todas tus conversaciones.

Este prompt reduce lo complaciente, no lo elimina. ChatGPT todavía se puede equivocar o darte la razón de vez en cuando. Si algo importa mucho, pídele que revise su propia respuesta.

Cómo usarlos

- Pega el prompt (o la versión completa) en Instrucciones personalizadas.
- Dale Guardar.
- Abre un chat nuevo y haz la pregunta trampa.
- Cuando tengas una idea o decisión, usa los prompts bonus.

Cuéntame qué te salió: coméntalo en el reel o etiquétame en @stephanobrooke.