# Cómo evitar que ChatGPT se invente cosas: 5 reglas y qué comprobar

> ChatGPT no consulta datos: **predice el texto más probable**, y por eso se inventa cosas con la misma seguridad con la que acierta. Para que te la cuele mucho menos: **dale la fuente, dale permiso para decir «no lo sé», haz que calcule con código, abre conversación nueva al cambiar de tema y comprueba nombres, cifras, citas y enlaces.** Debajo tienes qué escribir exactamente en cada caso.

## Las 5 reglas, con lo que tienes que escribir

| Qué escribes | Qué hace | Cuándo usarlo |
|---|---|---|
| **1. Dale la fuente.**<br>Pega el documento y añade: *«Responde solo con lo que pone en este texto. Si no está, dímelo.»* | El texto entra en su ventana de contexto: lo lee en vez de tirar de lo que recuerda, que está comprimido y es aproximado. | Siempre que el dato lo tengas tú: un contrato, una normativa, la documentación de una librería, las notas de una reunión. |
| **2. Dale permiso para no saber.**<br>*«Si no estás seguro, dímelo. Prefiero un "no lo sé" a una respuesta inventada.»* | Le quita la presión de contestar a toda costa, que es lo que su entrenamiento premia. | En preguntas de dato concreto: fechas, cifras, quién dijo qué, la biografía de alguien poco conocido. |
| **3. Para contar y calcular, código.**<br>*«Hazlo con Python y enséñame el resultado.»* | El código ve cada letra y cada cifra; el modelo solo ve trozos de texto. | Contar letras o palabras, sumar, sacar porcentajes, ordenar una lista, calcular días entre fechas. |
| **4. Tema nuevo, conversación nueva.**<br>Abre otra y pega: *«Contexto: [tres líneas con lo importante]»* | Empieza con la ventana limpia: lo importante no compite con veinte mensajes viejos por su atención. | Cuando cambias de tema, o cuando empieza a ignorar instrucciones que le diste al principio. |
| **5. Verifica lo que importa.**<br>*«Para cada dato, dame la fuente: enlace y frase exacta.»* | Te deja algo que comprobar. **No garantiza que la fuente exista**: también puede inventarse el enlace. | Antes de publicar, enviar o firmar algo. Abre cada enlace y busca la frase tú. |

## Qué comprobar siempre

| Lo que te da | Por qué ahí falla más | Cómo lo compruebas |
|---|---|---|
| Nombres de personas o empresas poco conocidas | Sabe perfectamente cómo suena una biografía y rellena los huecos con lo probable. El formato, perfecto; los datos, inventados. | Búscalo fuera del chat. Si no aparece en ninguna fuente, trátalo como inventado. |
| Cifras y fechas | Una cifra inventada suena exactamente igual de bien que una real. | Ve a la fuente original. Si es un cálculo, regla 3: que lo haga con código. |
| Citas textuales | Genera la frase que «suena» a esa persona, no la que dijo. | Busca la frase exacta entre comillas en un buscador. |
| Enlaces, artículos, sentencias | Produce títulos y URLs con forma creíble. En 2023 un abogado de Nueva York presentó seis sentencias sacadas de ChatGPT; ninguna existía y acabó sancionado. | Ábrelos todos. Un enlace sin abrir no es una fuente. |
| Cualquier cosa reciente | Tiene fecha de corte: lo que pasó después de su entrenamiento no lo sabe. | Pídele que busque en internet y que te enlace lo que encuentre, o compruébalo tú. |

## Por qué falla: lo que pasa por dentro

| Lo que pasa dentro | Cómo lo notas | Qué haces |
|---|---|---|
| **Predice el siguiente trozo de texto**, el más probable, no el verdadero. | Te contesta con el mismo tono cuando acierta que cuando se lo inventa. | Reglas 2 y 5: dale permiso para no saber y comprueba lo que importa. |
| **Ve tokens, no letras.** «Ferrocarril» le llega como `fer · roc · arr · il`; «123456789», como tres bloques de tres cifras. | Falla contando letras o multiplicando números largos. | Regla 3: código. |
| **Lo que sabe está comprimido** en sus parámetros (GPT-3 ya tenía 175.000 millones), como patrones, no como datos guardados. | Recuerda el estilo de las cosas, pero no el dato exacto. | Regla 1: si el dato existe en un documento, dáselo. |
| **No tiene memoria entre mensajes.** Cada vez se le reenvía toda la conversación desde el principio. | Parece que «se olvida» de lo que le dijiste. | Regla 4. Y lo importante, repítelo cerca del final. |

| Lo que pasa dentro | Cómo lo notas | Qué haces |
|---|---|---|
| **La ventana de contexto se acaba**, y lo que queda en medio se aprovecha peor que lo del principio y el final. | En conversaciones larguísimas ignora instrucciones del principio. | Conversación nueva, o las instrucciones clave al principio y al final. |
| **Se cobra y se limita por tokens**, y el español gasta algo más: «La inteligencia artificial no sabe nada.» son 7 tokens; en inglés, 5. | Llegas antes al límite de la conversación o del plan. | Pega solo la parte del documento que hace falta, no el documento entero. |

## Errores típicos

- **Preguntarle «¿seguro?» y fiarte de lo que diga.** No tiene dónde comprobarlo: te dará otra respuesta probable, que puede estar igual de inventada. Si quieres corregirlo, dale el dato o la fuente.
- **Dar por buena una fuente porque tiene buena pinta.** Un título de artículo bien formado y una URL con dominio real se generan igual que el resto del texto. Hasta que no lo abres, no existe.
- **Pedirle cuentas de cabeza.** Sumar una columna, contar caracteres o calcular un porcentaje «a ojo» es justo lo que peor hace. Con una línea de código no falla.
- **Arrastrar la misma conversación durante días.** Las instrucciones del principio acaban en medio de un contexto enorme y dejan de pesar. Abre otra y pega un resumen.
- **Creer que con el modelo más nuevo ya no se inventa nada.** Se equivoca bastante menos y puede buscar en internet, pero la alucinación no se arregla con un parche: se reduce. Lo que vayas a firmar, compruébalo.

## Preguntas frecuentes

### ¿Por qué ChatGPT se inventa cosas?

Porque no consulta ninguna base de datos: predice el texto más probable a partir de lo que vio al entrenarse. Casi siempre lo probable coincide con lo verdadero, pero cuando no sabe algo genera lo que suena bien, con la misma seguridad. Además, la forma en que se entrenan y se evalúan los modelos premia contestar antes que decir «no lo sé».

### ¿Se puede conseguir que ChatGPT no alucine nunca?

No. Se reduce mucho: los modelos nuevos se equivocan bastante menos, pueden buscar en internet y las cinco reglas de esta chuleta cortan la mayoría de los casos. Pero nace de cómo funciona por dentro, así que lo que importa (nombres, cifras, citas y enlaces) se sigue comprobando.

### ¿Por qué ChatGPT no sabe contar las letras de una palabra?

Porque no ve letras, ve tokens: trozos de texto convertidos en números. «Ferrocarril» le llega como cuatro trozos (fer, roc, arr, il) y la primera doble erre queda partida entre dos. Para contar o calcular, pídele que lo haga con código: el código sí ve cada letra y cada cifra.

## ¿Por qué ChatGPT se olvida de lo que le dije al principio?

Porque solo puede mirar lo que cabe en su ventana de contexto. Cuando la conversación se pasa, lo más antiguo se recorta o se resume, y aunque quepa todo, los modelos aprovechan peor lo que queda en medio. Si cambias de tema, abre una conversación nueva con un resumen corto.

## ¿ChatGPT tiene memoria?

No entre un mensaje y otro: cada vez que escribes, se le vuelve a mandar toda la conversación desde el principio. La «memoria» que recuerda a qué te dedicas son notas que se meten en ese mismo contexto antes de tu mensaje, no un recuerdo de verdad.