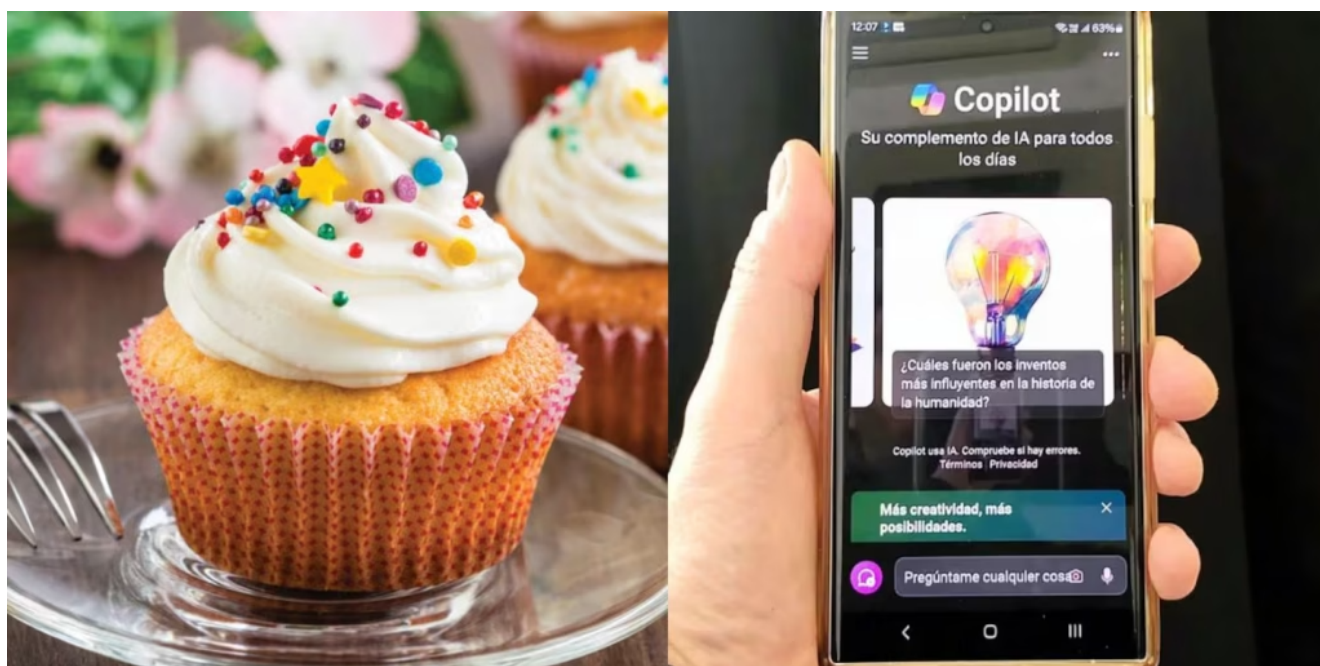


El truco del “Cupcake” para detectar cuando una Inteligencia Artificial inventa respuestas

31/03/2026



El llamado “truco Cupcake” se volvió viral entre periodistas y expertos en IA como una táctica simple para detectar cuando los modelos se inventan cosas. Consiste en **pedir detalles concretos sobre un elemento**, por ejemplo, un cupcake en una imagen, y contrastar la respuesta con lo que realmente puede verificarse en el archivo visual o la fuente.

En qué consiste la prueba del “Cupcake”,

Modelos populares como **ChatGPT**, **Gemini** y **Copilot** suelen ofrecer explicaciones detalladas aun cuando carecen de datos precisos, una conducta que la comunidad denomina “**hallucination**”. El truco funciona porque **obliga al**

sistema a entrar en pormenores verificables: ingredientes, etiquetas, marcas o texto visible. Si el modelo inventa elementos que no están, está fabricando información.

Para aplicar el truco Cupcake, se sugiere **pedir datos concretos: listar ingredientes exactos, cantidades o textos presentes en la foto.** Si la IA proporciona rótulos o números que no se ven, se detecta la invención. Es una técnica práctica que ayuda a **periodistas y usuarios a poner a prueba la fiabilidad de respuestas automatizadas.**

Ventajas y límites del truco del “Cupcake”

No obstante, el método tiene límites: **si la imagen es borrosa o la información es ambigua, la IA puede equivocarse sin “inventar” con mala intención.** Por eso los verificadores deben combinar el truco Cupcake con otras comprobaciones: fuentes externas, metadatos y sentido común periodístico para **evitar falsos positivos y no despreciar información correcta generada por el modelo.**

Redacciones y usuarios curiosos están probando la maniobra porque **no exige conocimientos técnicos y puede revelar fallas en segundos.** Basta una pregunta dirigida y una comparación visual o documental. El objetivo no es desacreditar la IA, sino **mejorar la verificación: entender cuándo conviene confiar y cuándo exigir fuentes y certezas adicionales.**

Los profesionales recomiendan documentar las interacciones: **guardar capturas, transcribir respuestas y pedir aclaraciones cuando haya contradicciones.** Interrogar varias veces con formulaciones distintas suele mostrar consistencia o inconsistencias. Además, como los modelos se actualizan, **una respuesta válida hoy puede cambiar mañana; por eso conviene archivar la evidencia durante la investigación periodística.**

Fuente: La 100