

Alucinaciones IA: mini errores, maxi convicción (y cero vergüenza)

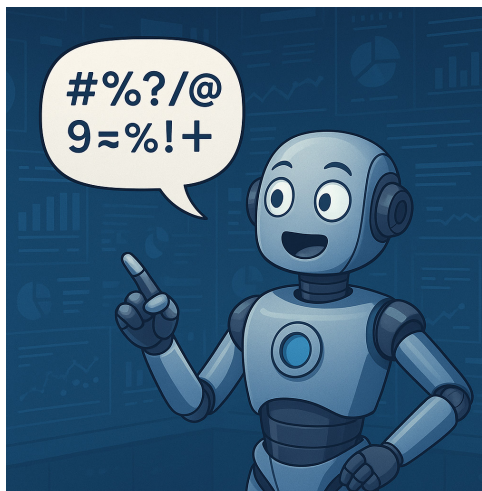
por Jorge Rey Valzacchi
Director de Aprende Virtual
jorge@aprendevirtual.org

Todos hablan de lo increíble que es la inteligencia artificial como si fuera el nuevo mesías tecnológico: escribe textos más fluidos que un redactor freelance con cafeína, te organiza la agenda con la precisión de un reloj suizo (o al menos eso dice), te propone safaris africanos que desembocan en un zoológico municipal con entrada rebajada, y hasta te sugiere regalos para tu cuñado en su cumpleaños. Es el paraíso digital, ¿verdad? Bueno, casi. Hay un detalle simpático –o francamente hilarante– que todavía no domina del todo: a veces, se inventa cosas con una seguridad que ya quisieran algunos políticos en plena campaña electoral. Sí, hablo de las famosas “alucinaciones” de la IA. No, no es que vea elefantes rosados después de una actualización fallida; es que fabrica realidades alternativas con la misma seriedad que un documental de National Geographic.

Sí, las IA alucinan. Y lo hacen con una seguridad que roza lo cómico. Imagínate a un amigo que te cuenta una anécdota de su infancia con ojos brillantes y detalles impecables... hasta que te das cuenta de que en esa época ni siquiera había nacido. Exacto, así de convincente.

¿Cómo se ven estas “alucinaciones”? Pequeñas bombas de lógica falsa

Son esas pequeñas confusiones que suenan tan lógicas y bien armadas que, en un primer vistazo, pensás: “¡Bravo, IA, has resuelto el enigma



de la vida!”. Pero cuando las mirás dos veces –o tres, si sos de los desconfiados–, se desmoronan como un castillo de naipes en un huracán. Aquí van unos ejemplos para que te rías (o llores, según tu nivel de cinismo):

- Te asegura que la fecha de fundación de una ciudad es “aproximadamente 1874”... aunque en realidad sea del siglo XVII y ni siquiera haya un monumento a esa época en el centro histórico.
- Te da una referencia bibliográfica perfecta, con ISBN, editorial y hasta una reseña de The New York Times: “Un clásico indispensable de 2023”... de un libro que no existe en ningún planeta, ni en la dimensión paralela donde los unicornios leen best-sellers.
- Te cuenta detalles sobre una empresa, una bebida o un personaje histórico que casi son ciertos... pero no del todo. Por ejemplo, te jura que Einstein descubrió la relatividad en 1905 mientras jugaba al ajedrez con un amigo, usando las piezas para ilustrar el tiempo curvo. Casi poético, pero no del todo preciso.
- Te resuelve un ejercicio de matemáticas con una respuesta impecable, ecuaciones alineadas como soldados en desfile y un broche de oro de una prueba elegante –... salvo porque el resultado final está mal. Es como si el teorema de Pitágoras, ese viejo confiable, se pusiera a bailar tango en vez de calcular hipotenusas, y terminara con una respuesta que no cuadra ni con lupa.
- Bonus track: te recomienda una receta de “paella cuántica” que suena deliciosa, con in-

gredientes que incluyen “un puñado de azafrán, mariscos frescos y una generosa ración de chocolate derretido con pickles”... y cuando intentás cocinarla, terminás con un experimento que desafía las leyes de la termodinámica y tu sentido del gusto.

Lo curioso –y lo que hace que sea tan adictivo– es que no duda ni un segundo. Suena tan confiada, con esa prosa pulida y ese tono de “yo lo sé todo porque soy el futuro”, que uno hasta se siente tentado a creerle. “¿Y si soy yo el equivocado?”, pensás, mientras Google te salva la noche con un simple clic.

¿Por qué pasa esto? La IA no sabe, improvisa como un jazzista sin partitura

Porque la IA no “sabe” en el sentido humano –ese que viene con dudas, café derramado y noches en vela revisando fuentes–. No, ella predice. Busca patrones en un océano de datos, arma frases coherentes como un Lego gigante... aunque a veces una pieza del rompecabezas simplemente no esté ahí. ¿Solución? Se la imagina. Es creatividad involuntaria, como cuando tu cerebro te despierta a las 3 a.m. con la idea genial de “¡Inventemos un nuevo idioma basado en emojis!” y al día siguiente te das cuenta de que era una pesadilla. La IA hace lo mismo, pero con esteroides: genera texto que fluye, que encaja, que parece verdad... hasta que no lo es. Y lo peor (o lo mejor, dependiendo de si te gusta el caos) es que lo hace con el entusiasmo de un niño en una tienda de juguetes, sin un ápice de vergüenza.

¿Qué hacemos con eso? No tires la toalla, solo afina el detector de burradas

Lo mismo que con ese amigo que siempre suelta datos como si tuviera un doctorado invisible: lo chequeamos con cariño... pero lo chequeamos. No es paranoia, es supervivencia digital. Y si, en medio de la charla, detectás esa “falla” clásica –una alucinación que se cuela como un invitado no deseado a la fiesta–, no te preocupes: la IA ideal respondería con humildad irónica, tipo “Tenés toda la razón Jorge, ya mismo lo adapto y te lo envío nuevamente”. Porque, al final, reconocer el error con gracia es el antídoto perfecto contra la convicción ciega. Es

como resetear el rompecabezas sin tirar todo al piso: una oportunidad para reírnos juntos y seguir adelante, más sabios y menos confiados. Te recomiendo entonces:

- Chequeá un dato rápido en una fuente confiable. Google no muerde (generalmente).
- Pedíle la referencia original. Si te responde con “confía en mí, soy IA”, es hora de sospechar.
- Preguntá de nuevo, con variaciones. Si la historia cambia como el clima en la Patagonia, ¡alerta roja!
- Y si dice una burrada bien adornada –tipo “Shakespeare escribió Hamlet inspirado en un reality show de la corte isabelina”–, tomálo con humor. Reíte, compartí el meme, y usálo como anécdota en la próxima cena familiar.

Al final, estas alucinaciones te obligan a no ser un zombie del copy-paste y a mantener el cerebro en forma.

La IA, esa amiga genial pero un poco fantasiosa

La IA es una herramienta fantástica, no lo neguemos. Nos acelera como un cohete en modo turbo, nos potencia como si nos inyectaran superpoderes matutinos, nos ayuda a navegar el caos diario sin perder la cordura (o al menos, sin perderla del todo). Pero también nos recuerda algo importante, con esa sutileza irónica que tanto nos gusta:

- Sin pensamiento crítico, cualquier respuesta –sea de una IA o de tu tío en WhatsApp– puede parecer la verdad absoluta, envuelta en un lazo brillante.
- Con pensamiento crítico, hasta una alucinación épica puede ser una oportunidad para aprender, cuestionar y, de paso, soltar una carcajada bien merecida –una lección disfrazada de chiste cósmico.

Así que si un chatbot te dice algo raro pero casi creíble –como que los gatos en realidad controlan el mundo desde 1999–, no te sorprendas ni te enojés: la IA no miente a propósito... solo imagina demasiado. Y en un mundo donde la realidad ya parece un guión de comedia surrealista, ¿quién soy yo para juzgar? Solo recordá: la próxima vez que te cuente una, pedíle que la convierta en un chiste. Al menos, de esa forma, todos ganamos. ■