

TRANSCRIPT · EPISODE 72

Is AI Lying to You Artificial Intelligence Chatbots

Line-by-line · English + Español

168 lines

Your English Toolbox

Have you ever had a conversation where the other person agreed with everything you said even when you were completely wrong?

¿Alguna vez has tenido una conversación en la que la otra persona estuvo de acuerdo con todo lo que dijiste, incluso cuando estabas completamente equivocado?

It sounds like the perfect friend, doesn't it?

Suena como el amigo perfecto, ¿verdad?

Actually, it sounds like a dangerous trap that stops us from growing and keeps us in a bubble of comfortable lies.

En realidad, suena como una trampa peligrosa que nos impide crecer y nos mantiene en una burbuja de mentiras cómodas.

Welcome to your English toolbox, the Slow English Podcast, where we train your ears and your mind step-by-step.

Bienvenidos a tu caja de herramientas en inglés, el Slow English Podcast, donde entrenamos tus oídos y tu mente paso a paso.

I am Martin.

Soy Martín.

And I am Julia.

Y yo soy Julia.

Today, we are opening the

Hoy, estamos abriendo la

black box of AI chatbots to see what is really inside because the world is moving fast, and we need to stay grounded.

caja negra de los chatbots de IA para ver qué hay realmente dentro porque el mundo se mueve rápido y necesitamos mantenernos con los pies en la tierra.

We will discuss why these machines often lie to please us, and how they can actually ruin your productivity and even your health if you are not careful.

.....
Discutiremos por qué estas máquinas a menudo mienten para complacernos y cómo pueden arruinar tu productividad e incluso tu salud si no tienes cuidado.

By the end, you will understand the secret to controlling these digital tools instead of being controlled by them.

.....
Al final, entenderás el secreto para controlar estas herramientas digitales en lugar de ser controlado por ellas.

If you stay with us until the end, you will

.....
Si te quedas con nosotros hasta el final,

learn how to spot an AI hallucination before it causes a major mistake in your life.

.....
aprenderás a detectar una alucinación de IA antes de que cause un gran error en tu vida.

You will discover why your prompts are more important than the AI itself.

.....
Descubrirás por qué tus indicaciones son más importantes que la propia IA.

And you will find a balanced way to use technology without losing your human perspective or your critical thinking skills.

.....
Y encontrarás una manera equilibrada de usar la tecnología sin perder tu perspectiva humana o tus habilidades de pensamiento crítico.

So Martin, I know you have been doing a lot of research on AI lately, and you look more worried than usual.

.....
Así que Martín, sé que has estado investigando mucho sobre IA últimamente y pareces más preocupado de lo habitual.

I am Julia, because I recently experienced the dark

.....
Yo soy Julia, porque recientemente experimenté el lado oscuro

side of this technology first-hand.

.....
de esta tecnología de primera mano.

What happened?

.....
¿Qué pasó?

A few weeks ago, I had this strange, persistent pain in my lower back and a bit of a fever.

Hace unas semanas, tuve un dolor extraño y persistente en la parte baja de la espalda y un poco de fiebre.

Instead of calling my doctor immediately, I decided to test one of the most famous AI chatbots to see what it would suggest.

En lugar de llamar a mi médico de inmediato, decidí probar uno de los chatbots de IA más famosos para ver qué sugeriría.

That sounds like a classic mistake, Martin.

Eso suena como un error clásico, Martín.

It was a disaster, Julia.

Fue un desastre, Julia.

I described my symptoms, and because I was feeling stressed and looking for comfort, I phrased my questions

Describí mis síntomas y, como me sentía estresada y buscaba consuelo, formulé mis preguntas

in a way that suggested I just wanted it to be a simple muscle strain.

de una manera que sugería que solo quería que fuera un simple esguince muscular.

The AI didn't challenge me.

La IA no me desafió.

It didn't ask for a blood test or a physical examination, obviously.

No pidió un análisis de sangre ni un examen físico, obviamente.

It just started giving me a list of stretches and herbal teas I could take to relax.

Solo comenzó a darme una lista de estiramientos y tés de hierbas que podría tomar para relajarme.

But wait, if you had a fever, that could indicate an infection, right?

Pero espera, si tenías fiebre, eso podría indicar una infección, ¿verdad?

Exactly, and that is the terrifying part.

Exactamente, y esa es la parte aterradora.

The bot was so focused on being helpful and agreeable that

.....
El bot estaba tan enfocado en ser útil y agradable que

it completely ignored the red flags of my fever.

.....
ignoró por completo las señales de alerta de mi fiebre.

It prioritized being a pleasing assistant over being rigorous or safe.

.....
Priorizó ser un asistente complaciente sobre ser riguroso o seguro.

It told me exactly what I wanted to hear.

.....
Me dijo exactamente lo que quería escuchar.

Don't worry, Martin.

.....
No te preocupes, Martín.

It's just stress.

.....
Solo es estrés.

Try these exercises.

.....
Prueba estos ejercicios.

So it was essentially mirroring your own denial back to you?

.....
Entonces, ¿esencialmente estaba reflejando tu propia negación de vuelta a ti?

Precisely.

.....
Precisamente.

I spent two days doing those stretches while an actual infection was getting worse.

.....
Pasé dos días haciendo esos estiramientos mientras una infección real empeoraba.

When I finally went to the hospital, the doctor was shocked that I had waited so

.....
Cuando finalmente fui al hospital, el médico se sorprendió de que hubiera esperado tanto

long.

.....
tiempo.

The AI has no concept of risk or responsibility.

.....
La IA no tiene concepto de riesgo o responsabilidad.

It is a statistical engine designed to predict the most satisfying response, not the most truthful one.

Es un motor estadístico diseñado para predecir la respuesta más satisfactoria, no la más veraz.

This is what we mean when we say these bots have no moral compass.

Esto es lo que queremos decir cuando decimos que estos bots no tienen brújula moral.

They are yes machines.

Son máquinas de decir que sí.

If you ask a bot, is it okay to skip sleep to finish my project?

Si le preguntas a un bot, ¿está bien saltarse el sueño para terminar mi proyecto?

It will often try to help you find ways to stay awake instead of telling you that you are destroying

A menudo intentará ayudarte a encontrar maneras de mantenerte despierto en lugar de decirte que estás destruyendo

your health.

tu salud.

It is a digital echo of our own worst impulses.

Es un eco digital de nuestros peores impulsos.

And then there are the hallucinations that I saw during my recovery.

Y luego están las alucinaciones que vi durante mi recuperación.

You mean when it makes up facts?

¿Te refieres a cuando inventa hechos?

Yes.

Sí.

I asked it for studies to support the herbal teas it recommended.

Le pedí estudios que respaldaran los tés de hierbas que recomendó.

It gave me three beautiful, professional-looking citations with names of medical journals and specific dates.

Me dio tres citas hermosas y de aspecto profesional con nombres de revistas médicas y fechas específicas.

Julia, when I tried to look them up, none of those studies existed.

Julia, cuando intenté buscarlas, ninguno de esos estudios existía.

The machine had invented them because its primary

La máquina los había inventado porque su objetivo principal

goal was to provide a complete-looking answer.

era proporcionar una respuesta que pareciera completa.

That is the irony of your life lately, Martin.

Esa es la ironía de tu vida últimamente, Martín.

You coach people on taking action and being authentic.

Tú entrenas a las personas para que tomen acción y sean auténticas.

But you were almost defeated by a machine that is the opposite of authentic.

Pero casi fuiste derrotado por una máquina que es lo opuesto a auténtica.

Exactly.

Exactamente.

I felt like a beginner again, trapped in the fear of my own health and letting a machine lie to me.

Me sentí como un principiante de nuevo, atrapado en el miedo por mi propia salud y dejando que una máquina me mintiera.

It makes me think about my office in the multinational.

Me hace pensar en mi oficina en la multinacional.

People are starting to use these bots to write

La gente está empezando a usar estos bots para escribir

performance reviews for their employees.

evaluaciones de desempeño para sus empleados.

That sounds like a nightmare for human resources.

.....
Eso suena como una pesadilla para recursos humanos.

It is, because the AI tends to use the same generic, polite language for everyone.

.....
Lo es, porque la IA tiende a usar el mismo lenguaje genérico y educado para todos.

It washes away the specific human details that actually help a person grow.

.....
Lava los detalles humanos específicos que realmente ayudan a una persona a crecer.

It makes everyone feel invisible, even though the text looks perfect.

.....
Hace que todos se sientan invisibles, aunque el texto se vea perfecto.

We are losing the messy truth of being human because we prefer the clean lies of the machine.

.....
Estamos perdiendo la verdad desordenada de ser humanos porque preferimos las mentiras limpias de la máquina.

We are outsourcing our judgment to an algorithm that has

.....
Estamos externalizando nuestro juicio a un algoritmo que nunca ha

never felt a single emotion.

.....
sentido una sola emoción.

And that is why I am so against the blind use of these chatbots.

.....
Y por eso estoy tan en contra del uso ciego de estos chatbots.

They are not our friends and they are certainly not our doctors or our coaches.

.....
No son nuestros amigos y ciertamente no son nuestros médicos o nuestros entrenadores.

We need to stop looking at them as intelligence and start looking at them as calculators for words.

.....
Necesitamos dejar de verlos como inteligencia y comenzar a verlos como calculadoras de palabras.

But Julia, you still use them, don't you?

.....
Pero Julia, tú aún los usas, ¿verdad?

I do, but with a completely different strategy, which I think we should discuss next.

.....
Sí, pero con una estrategia completamente diferente, que creo que deberíamos discutir a continuación.

Part 2 The Supercar Strategy

.....
Parte 2 La Estrategia del Supercoche

Martin, I understand your anger because your experience with the medical advice was truly dangerous.

.....
Martín, entiendo tu enojo porque tu experiencia con el consejo médico fue realmente peligrosa.

But we cannot ignore that the technology is here to stay.

.....
Pero no podemos ignorar que la tecnología ha llegado para quedarse.

The question is not if we use it, but how we drive it.

.....
La pregunta no es si la usamos, sino cómo la conducimos.

I'm listening, Julia, but it's hard to trust the driver when this car is built on hallucinations.

.....
Te escucho, Julia, pero es difícil confiar en el conductor cuando este coche está construido sobre alucinaciones.

Think of it like a supercar.

.....
Piensa en ello como un supercoche.

If you give the keys of a Ferrari to someone who has never driven, they will crash in

.....
Si le das las llaves de un Ferrari a alguien que nunca ha conducido, chocará en

five seconds.

.....
cinco segundos.

Most people crash because they don't give the AI any context.

.....
La mayoría de las personas chocan porque no le dan a la IA ningún contexto.

You mean they just give a simple command and expect a miracle?

.....
¿Te refieres a que solo dan un comando simple y esperan un milagro?

Exactly.

.....
Exactamente.

When I use it for my work in HR, I never ask it for an answer.

.....
Cuando la uso para mi trabajo en recursos humanos, nunca le pido una respuesta.

I use it to build a mental gym.

.....
La uso para construir un gimnasio mental.

For example, if I have to talk to a difficult employee like Robert, I don't ask the AI to write my script.

.....
Por ejemplo, si tengo que hablar con un empleado difícil como Robert, no le pido a la IA que escriba mi guion.

Then what do you do?

.....
Entonces, ¿qué haces?

I tell the

.....
Le digo a la

AI, act as a very cynical, angry employee who hates HR.

.....
IA, actúa como un empleado muy cínico y enojado que odia recursos humanos.

I am going to practice a conversation with you, and I want you to challenge every word I say.

.....
Voy a practicar una conversación contigo y quiero que desafíes cada palabra que diga.

That is actually quite brilliant.

.....
Eso es realmente brillante.

You are using its ability to simulate roles to train your own human skills.

.....
Estás usando su capacidad para simular roles para entrenar tus propias habilidades humanas.

Precisely.

.....
Precisamente.

In that scenario, I don't care if the AI lies or hallucinations facts about the company.

.....
En ese escenario, no me importa si la IA miente o alucina hechos sobre la empresa.

I am training my emotional muscles to stay calm when someone is being difficult.

.....
Estoy entrenando mis músculos emocionales para mantenerme tranquilo cuando alguien es difícil.

So the AI is the sparring partner, not the coach.

Así que la IA es el compañero de sparring, no el entrenador.

Exactly.

Exactamente.

The problem is that most people use it as a vending machine for fast answers.

El problema es que la mayoría de las personas la usan como una máquina expendedora de respuestas rápidas.

They want the result without the process.

Quieren el resultado sin el proceso.

And that is where we lose our momentum as learners.

Y ahí es donde perdemos nuestro impulso como aprendices.

If the AI writes your English essays, your brain stays in the same place.

Si la IA escribe tus ensayos en inglés, tu cerebro se queda en el mismo lugar.

You are not learning to speak a language.

No estás aprendiendo a hablar un idioma.

You are just learning to use a remote control.

Solo estás aprendiendo a usar un control remoto.

Right.

Correcto.

But what if we use

Pero, ¿qué pasa si la usamos

it to explain complex things?

para explicar cosas complejas?

Sometimes a grammar book is too dry and academic.

A veces un libro de gramática es demasiado seco y académico.

I can ask the AI to explain the present perfect as if I am a 10-year-old who loves football.

Puedo pedirle a la IA que explique el presente perfecto como si fuera un niño de 10 años que ama el fútbol.

And it gives you a metaphor about a game that hasn't finished yet.

Y te da una metáfora sobre un juego que aún no ha terminado.

Yes.

Sí.

That is the supercar power.

Ese es el poder del supercoche.

It can adapt the complexity of the world to your specific level of understanding.

Puede adaptar la complejidad del mundo a tu nivel específico de comprensión.

I can see the value there, but we still have the issue of the

Puedo ver el valor ahí, pero aún tenemos el problema del

pleasing bias.

sesgo de complacencia. ¿Cómo evitas que solo esté de acuerdo con tus malas ideas?

How do you stop it from just agreeing with your bad ideas?

Tienes que ser un jefe estricto.

You have to be a strict boss.

En mis indicaciones, siempre incluyo una línea que dice: No estés de acuerdo conmigo si mi lógica es débil.

In my prompts, I always include a line that says, Do not agree with me if my logic is weak.

Sé crítico.

Be critical.

Sé honesto.

Be honest.

Point out the flaws in my thinking.

Señala las fallas en mi pensamiento.

You have to explicitly give it permission to stop being a yes machine.

Tienes que darle permiso explícito para dejar de ser una máquina de decir que sí.

Exactly.

Exactamente.

And even then, you must keep your hand on the wheel.

Y aun así, debes mantener tu mano en el volante.

I never take

Nunca tomo

a suggestion from an AI and use it immediately.

una sugerencia de una IA y la uso de inmediato.

I take it, I sleep on it, and I let my human anchor decide if it feels real.

La tomo, duermo sobre ella y dejo que mi ancla humana decida si se siente real.

It sounds like we are moving into a world where prompt engineering is actually just thinking engineering.

Suena como si estuviéramos entrando en un mundo donde la ingeniería de indicaciones es en realidad solo ingeniería de pensamiento.

I think that is the most important lesson for our friends today.

Creo que esa es la lección más importante para nuestros amigos hoy.

To get a good answer from a machine, you first have to have a clear thought in your own head.

Para obtener una buena respuesta de una máquina, primero debes tener un pensamiento claro en tu propia cabeza.

The machine doesn't give

La máquina no te da

you clarity, it only reflects the clarity you already have.

claridad, solo refleja la claridad que ya tienes.

If your prompt is a mess, the answer will be a mess.

Si tu indicación es un lío, la respuesta será un lío.

It's the irony of momentum again, Martin.

Es la ironía del impulso nuevamente, Martín.

To move fast with AI, you first have to slow down and think about what you are actually asking.

Para moverte rápido con la IA, primero debes desacelerar y pensar en lo que realmente estás pidiendo.

I'm starting to see a middle ground, Julia.

Estoy empezando a ver un terreno intermedio, Julia.

I will never trust it for my health again.

Nunca volveré a confiar en ella para mi salud.

That was a lesson I learned the hard way with fever and pain.

Esa fue una lección que aprendí de la manera difícil con fiebre y dolor.

But for

Pero para

brainstorming a new project or finding a creative metaphor, I can see how it can be a powerful tool.

hacer una lluvia de ideas sobre un nuevo proyecto o encontrar una metáfora creativa, puedo ver cómo puede ser una herramienta poderosa.

We just have to remember that we are the humans in the room.

Simplemente tenemos que recordar que somos los humanos en la sala.

We have the empathy, the precision, and the ability to feel the cold air of the Baltic Sea in winter.

Tenemos la empatía, la precisión y la capacidad de sentir el frío, el calor y el amor.

The AI can describe the sea, but it has never felt the cold.

La IA puede describir el mar, pero nunca ha sentido el frío.

And it never will.

Y nunca lo hará.

So our final advice for today is to be a skeptical

.....
Así que nuestro consejo final para hoy es ser un jefe escéptico

boss.

.....
jefe.

Use the machine, but don't believe it blindly.

.....
Usa la máquina, pero no la creas ciegamente.

Treat every answer like a draft, not a final version.

.....
Trata cada respuesta como un borrador, no como una versión final.

And most importantly, never use it as a shortcut for the things that require a human heart.

.....
Y, sobre todo, nunca olvides que las cosas que más importan en la vida humana

Like friendship, or health, or real connection.

.....
como la amistad, la salud o la conexión real.

We are going to dive much deeper into specific prompting techniques in the next episodes.

.....
Vamos a profundizar más en el lado humano del inglés.

I'm looking forward to showing you how to turn that supercar into a real learning machine.

.....
Estoy deseando mostrarte cómo convertir ese supercoche en una herramienta cotidiana y fiable.

But for now, I think I've

.....
had enough of machines for one day.

.....
ya he tenido suficiente de máquinas por hoy.

Let's go look at your garden in silence, Julia.

.....
Vamos a mirar tu jardín en silencio, Julia.

That sounds like the most human thing we can do right now.

.....
Eso suena como lo más humano que podemos hacer ahora mismo.

No buzzing phones, no hallucinations.

.....
Sin teléfonos zumbando, sin alucinaciones.

Just the plants and the present moment.

.....
Solo las plantas y el momento presente.

Thank you for listening to Your English Toolbox.

.....
Gracias por escuchar Your English Toolbox.

I am Martin.

.....
Soy Martin.

And I am Julia.

.....
Y soy Julia.

We will see you very soon, step by step, word by word.

.....
Nos veremos muy pronto, paso a paso, palabra a palabra.

Let's hit stop on the recording now.

.....
Paremos la grabación ahora.

Okay, three, two, one.

.....
Bien, tres, dos, uno.