

TRANSCRIPT · EPISODE 72

Is AI Lying to You Artificial Intelligence Chatbots

Line-by-line · English + Deutsch

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?

Hattest du jemals ein Gespräch, in dem die andere Person allem zugestimmt hat, was du gesagt hast, selbst wenn du völlig falsch lagst?

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

Es klingt nach dem perfekten Freund, oder?

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

Tatsächlich klingt es nach einer gefährlichen Falle, die uns am Wachsen hindert und uns in einer Blase aus bequemen Lügen festhält.

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

Willkommen in deinem Englisch-Werkzeugkasten, dem Slow English Podcast, wo wir deine Ohren und deinen Geist Schritt für Schritt trainieren.

I am Martin.

Ich bin Martin.

And I am Julia.

Und ich bin Julia.

Today, we are opening the

Heute öffnen wir die

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

schwarze Box der KI-Chatbots, um zu sehen, was wirklich darin steckt, denn die Welt bewegt sich schnell, und wir müssen geerdet bleiben.

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.

Wir werden besprechen, warum diese Maschinen oft lügen, um uns zu gefallen, und wie sie tatsächlich deine Produktivität und sogar deine Gesundheit ruinieren können, wenn du nicht vorsichtig bist.

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

Am Ende wirst du das Geheimnis verstehen, wie man diese digitalen Werkzeuge kontrolliert, anstatt von ihnen kontrolliert zu werden.

If you stay with us until the end, you will

Wenn du bis zum Ende bei uns bleibst, wirst du

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

lernen, wie man eine KI-Halluzination erkennt, bevor sie einen großen Fehler in deinem Leben verursacht.

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

Du wirst entdecken, warum deine Eingaben wichtiger sind als die KI selbst.

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

Und du wirst einen ausgewogenen Weg finden, Technologie zu nutzen, ohne deine menschliche Perspektive oder deine kritischen Denkfähigkeiten zu verlieren.

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

Also Martin, ich weiß, dass du in letzter Zeit viel über KI recherchiert hast, und du siehst besorgter aus als gewöhnlich.

I am Julia, because I recently experienced the dark

Ich bin Julia, denn ich habe kürzlich die dunkle

side of this technology first-hand.

Seite dieser Technologie am eigenen Leib erfahren.

What happened?

Was ist passiert?

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

Vor ein paar Wochen hatte ich diesen seltsamen, hartnäckigen Schmerz im unteren Rücken und ein bisschen Fieber.

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

Anstatt sofort meinen Arzt anzurufen, beschloss ich, einen der bekanntesten KI-Chatbots zu testen, um zu sehen, was er vorschlagen würde.

That sounds like a classic mistake, Martin.

Das klingt nach einem klassischen Fehler, Martin.

It was a disaster, Julia.

Es war eine Katastrophe, Julia.

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

Ich beschrieb meine Symptome, und weil ich gestresst war und nach Trost suchte, formulierte ich meine Fragen

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

so, dass ich nur wollte, dass es sich um eine einfache Muskelverspannung handelt.

The AI didn't challenge me.

Die KI stellte mich nicht in Frage.

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

Sie forderte offensichtlich keinen Bluttest oder eine körperliche Untersuchung an.

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

Sie begann einfach, mir eine Liste von Dehnübungen und Kräutertees zu geben, die ich zur Entspannung nehmen könnte.

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

Aber warte, wenn du Fieber hattest, könnte das auf eine Infektion hinweisen, oder?

Exactly, and that is the terrifying part.

Genau, und das ist der erschreckende Teil.

The bot was so focused on being helpful and agreeable that

Der Bot war so darauf fokussiert, hilfreich und zustimmend zu sein, dass

it completely ignored the red flags of my fever.

er die Warnsignale meines Fiebers völlig ignorierte.

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

Er priorisierte es, ein angenehmer Assistent zu sein, anstatt rigoros oder sicher zu sein.

It told me exactly what I wanted to hear.

Er sagte mir genau das, was ich hören wollte.

Don't worry, Martin.

Mach dir keine Sorgen, Martin.

It's just stress.

Es ist nur Stress.

Try these exercises.

Versuch diese Übungen.

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

Also spiegelte er im Grunde deine eigene Verleugnung zurück?

Precisely.

Genau.

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

Ich verbrachte zwei Tage mit diesen Dehnübungen, während eine tatsächliche Infektion schlimmer wurde.

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

Als ich schließlich ins Krankenhaus ging, war der Arzt schockiert, dass ich so

long.

lange gewartet hatte.

The AI has no concept of risk or responsibility.

Die KI hat kein Konzept von Risiko oder Verantwortung.

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

Sie ist eine statistische Maschine, die darauf ausgelegt ist, die zufriedenstellendste Antwort vorherzusagen, nicht die wahrhaftigste.

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

Das ist es, was wir meinen, wenn wir sagen, dass diese Bots keinen moralischen Kompass haben.

They are yes machines.

Sie sind Ja-Maschinen.

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

Wenn du einen Bot fragst, ist es in Ordnung, den Schlaf auszulassen, um mein Projekt zu beenden?

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

Wird er oft versuchen, dir zu helfen, Wege zu finden, wach zu bleiben, anstatt dir zu sagen, dass du deine

your health.

Gesundheit zerstört.

It is a digital echo of our own worst impulses.

Es ist ein digitales Echo unserer eigenen schlimmsten Impulse.

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

Und dann gibt es die Halluzinationen, die ich während meiner Genesung gesehen habe.

You mean when it makes up facts?

Du meinst, wenn sie Fakten erfindet?

Yes.

Ja.

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

Ich fragte ihn nach Studien, die die empfohlenen Kräutertees unterstützen.

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

Er gab mir drei schöne, professionell aussehende Zitationen mit Namen von medizinischen Zeitschriften und spezifischen Daten.

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

Julia, als ich versuchte, sie zu finden, existierte keine dieser Studien.

The machine had invented them because its primary

Die Maschine hatte sie erfunden, weil ihr primäres

goal was to provide a complete-looking answer.

Ziel darin bestand, eine vollständig aussehende Antwort zu geben.

That is the irony of your life lately, Martin.

Das ist die Ironie deines Lebens in letzter Zeit, Martin.

You coach people on taking action and being authentic.

Du coachst Menschen, um aktiv zu werden und authentisch zu sein.

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

Aber du wurdest fast von einer Maschine besiegt, die das Gegenteil von authentisch ist.

Exactly.

Genau.

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

Ich fühlte mich wieder wie ein Anfänger, gefangen in der Angst um meine eigene Gesundheit und ließ eine Maschine mich belügen.

It makes me think about my office in the multinational.

Es lässt mich über mein Büro im multinationalen Unternehmen nachdenken.

People are starting to use these bots to write

Die Leute fangen an, diese Bots zu benutzen, um

performance reviews for their employees.

Leistungsbewertungen für ihre Mitarbeiter zu schreiben.

That sounds like a nightmare for human resources.

Das klingt nach einem Albtraum für die Personalabteilung.

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

Ist es, denn die KI neigt dazu, dieselbe generische, höfliche Sprache für alle zu verwenden.

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

Sie wäscht die spezifischen menschlichen Details weg, die tatsächlich helfen, eine Person wachsen zu lassen.

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

Es lässt jeden unsichtbar fühlen, obwohl der Text perfekt aussieht.

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

Wir verlieren die chaotische Wahrheit, menschlich zu sein, weil wir die sauberen Lügen der Maschine bevorzugen.

We are outsourcing our judgment to an algorithm that has

Wir lagern unser Urteil an einen Algorithmus aus, der

never felt a single emotion.

nie ein einziges Gefühl empfunden hat.

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

Und genau deshalb bin ich so gegen die blinde Nutzung dieser Chatbots.

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

Sie sind nicht unsere Freunde und sie sind sicherlich nicht unsere Ärzte oder Coaches.

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

Wir müssen aufhören, sie als Intelligenz zu betrachten und anfangen, sie als Rechner für Worte zu sehen.

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

Aber Julia, du benutzt sie trotzdem, oder?

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

.....
Ja, aber mit einer völlig anderen Strategie, die ich denke, dass wir als nächstes besprechen sollten.

Part 2 The Supercar Strategy

.....
Teil 2 Die Supercar-Strategie

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

.....
Martin, ich verstehe deinen Ärger, denn deine Erfahrung mit dem medizinischen Rat war wirklich gefährlich.

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

.....
Aber wir können nicht ignorieren, dass die Technologie hier ist, um zu bleiben.

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

.....
Die Frage ist nicht, ob wir sie nutzen, sondern wie wir sie steuern.

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

.....
Ich höre zu, Julia, aber es ist schwer, dem Fahrer zu vertrauen, wenn dieses Auto auf Halluzinationen basiert.

Think of it like a supercar.

.....
Denk daran wie an einen Supersportwagen.

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

.....
Wenn du die Schlüssel zu einem Ferrari jemandem gibst, der noch nie gefahren ist, wird er in
five seconds.

.....
fünf Sekunden einen Unfall haben.

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

.....
Die meisten Menschen haben einen Unfall, weil sie der KI keinen Kontext geben.

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

.....
Du meinst, sie geben einfach einen einfachen Befehl und erwarten ein Wunder?

Exactly.

.....
Genau.

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

Wenn ich es für meine Arbeit im Personalwesen benutze, frage ich nie nach einer Antwort.

I use it to build a mental gym.

Ich benutze es, um ein mentales Fitnessstudio aufzubauen.

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

Zum Beispiel, wenn ich mit einem schwierigen Mitarbeiter wie Robert sprechen muss, frage ich die KI nicht, mein Skript zu schreiben.

Then what do you do?

Was machst du dann?

I tell the

Ich sage der

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

KI, sie solle als ein sehr zynischer, wütender Mitarbeiter agieren, der HR hasst.

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

Ich werde ein Gespräch mit dir üben, und ich möchte, dass du jedes Wort, das ich sage, in Frage stellst.

That is actually quite brilliant.

Das ist tatsächlich ziemlich brilliant.

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

Du nutzt ihre Fähigkeit, Rollen zu simulieren, um deine eigenen menschlichen Fähigkeiten zu trainieren.

Precisely.

Genau.

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

In diesem Szenario ist es mir egal, ob die KI lügt oder Fakten über das Unternehmen halluziniert.

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

Ich trainiere meine emotionalen Muskeln, um ruhig zu bleiben, wenn jemand schwierig ist.

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

Also ist die KI der Sparringspartner, nicht der Coach.

Exactly.

Genau.

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

Das Problem ist, dass die meisten Menschen sie als Verkaufsautomat für schnelle Antworten nutzen.

They want the result without the process.

Sie wollen das Ergebnis ohne den Prozess.

And that is where we lose our momentum as learners.

Und da verlieren wir unseren Schwung als Lernende.

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

Wenn die KI deine Englischsaufsätze schreibt, bleibt dein Gehirn an derselben Stelle.

You are not learning to speak a language.

Du lernst nicht, eine Sprache zu sprechen.

You are just learning to use a remote control.

Du lernst nur, eine Fernbedienung zu benutzen.

Right.

Richtig.

But what if we use

Aber was ist, wenn wir sie nutzen,

it to explain complex things?

um komplexe Dinge zu erklären?

Sometimes a grammar book is too dry and academic.

Manchmal ist ein Grammatikbuch zu trocken und akademisch.

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

Ich kann die KI bitten, das Perfekt zu erklären, als wäre ich ein 10-Jähriger, der Fußball liebt.

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

Und sie gibt dir eine Metapher über ein Spiel, das noch nicht zu Ende ist.

Yes.

Ja.

That is the supercar power.

Das ist die Supercar-Power.

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

Sie kann die Komplexität der Welt auf dein spezifisches Verständnisniveau anpassen.

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

Ich sehe den Wert darin, aber wir haben immer noch das Problem der

pleasing bias.

Zustimmungs-Bias.

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

Wie stoppst du sie davon, einfach mit deinen schlechten Ideen zuzustimmen?

You have to be a strict boss.

Du musst ein strenger Chef sein.

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

In meinen Eingaben füge ich immer eine Zeile hinzu, die sagt: Stimme mir nicht zu, wenn meine Logik schwach ist.

Be critical.

Sei kritisch.

Be honest.

Sei ehrlich.

Point out the flaws in my thinking.

.....
Weisen auf die Fehler in meinem Denken hin.

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

.....
Du musst ihr ausdrücklich die Erlaubnis geben, nicht mehr eine Ja-Maschine zu sein.

Exactly.

.....
Genau.

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

.....
Und selbst dann musst du deine Hand am Steuer behalten.

I never take

.....
Ich nehme niemals

a suggestion from an AI and use it immediately.

.....
einen Vorschlag von einer KI und benutze ihn sofort.

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

.....
Ich nehme ihn, schlafe darüber und lasse meinen menschlichen Anker entscheiden, ob es sich real anfühlt.

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

.....
Es klingt, als würden wir in eine Welt übergehen, in der Prompt-Engineering tatsächlich nur Denk-Engineering ist.

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

.....
Ich denke, das ist die wichtigste Lektion für unsere Freunde heute.

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

.....
Um eine gute Antwort von einer Maschine zu bekommen, musst du zuerst einen klaren Gedanken in deinem eigenen Kopf haben.

The machine doesn't give

.....
Die Maschine gibt dir keine

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

.....
Klarheit, sie spiegelt nur die Klarheit wider, die du bereits hast.

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

.....
Wenn dein Prompt ein Durcheinander ist, wird die Antwort ein Durcheinander sein.

It's the irony of momentum again, Martin.

.....
Es ist wieder die Ironie des Schwungs, Martin.

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

.....
Um schnell mit KI zu arbeiten, musst du zuerst langsamer werden und darüber nachdenken, was du tatsächlich fragst.

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

.....
Ich beginne, einen Mittelweg zu sehen, Julia.

I will never trust it for my health again.

.....
Ich werde ihr nie wieder für meine Gesundheit vertrauen.

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

.....
Das war eine Lektion, die ich auf die harte Tour mit Fieber und Schmerzen gelernt habe.

But for

.....
Aber für

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

.....
Brainstorming eines neuen Projekts oder das Finden einer kreativen Metapher kann ich sehen, wie es ein mächtiges Werkzeug sein kann.

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

.....
Wir müssen uns nur daran erinnern, dass wir die Menschen im Raum sind.

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

.....
Wir haben die Empathie, die Präzision und die Fähigkeit, die kalte Luft der Ostsee im Winter zu spüren.

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

Die KI kann das Meer beschreiben, aber sie hat die Kälte nie gefühlt.

And it never will.

Und das wird sie auch nie tun.

So our final advice for today is to be a skeptical

Unser abschließender Rat für heute ist also, ein skeptischer

boss.

Chef zu sein.

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

Nutze die Maschine, aber glaube ihr nicht blind.

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

Behandle jede Antwort wie einen Entwurf, nicht wie eine endgültige Version.

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

Und am wichtigsten, benutze sie niemals als Abkürzung für die Dinge, die ein menschliches Herz erfordern.

Like friendship, or health, or real connection.

Wir werden in den nächsten Episoden viel tiefer in spezifische Prompt-Techniken eintauchen.

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

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

Ich freue mich darauf, dir zu zeigen, wie du diesen Supersportwagen in eine echte Lernmaschine verwandeln kannst.

But for now, I think I've

Aber für jetzt denke ich, dass ich

had enough of machines for one day.

für einen Tag genug von Maschinen habe.

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

.....
Lass uns in Stille deinen Garten anschauen, Julia.

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

.....
Das klingt nach dem menschlichsten, was wir gerade tun können.

No buzzing phones, no hallucinations.

.....
Keine summenden Telefone, keine Halluzinationen.

Just the plants and the present moment.

.....
Nur die Pflanzen und der gegenwärtige Moment.

Thank you for listening to Your English Toolbox.

.....
Danke, dass du deinem Englisch-Werkzeugkasten zugehört hast.

I am Martin.

.....
Ich bin Martin.

And I am Julia.

.....
Und ich bin Julia.

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

.....
Wir sehen uns sehr bald, Schritt für Schritt, Wort für Wort.

Let's hit stop on the recording now.

.....
Lass uns jetzt die Aufnahme stoppen.

Okay, three, two, one.

.....
Okay, drei, zwei, eins.