

TRANSCRIPT · EPISODE 72

Is AI Lying to You Artificial Intelligence Chatbots

Full episode script · English

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?

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

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

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

I am Martin.

And I am Julia.

Today, we are opening the

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

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.

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

If you stay with us until the end, you will

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

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

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

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

I am Julia, because I recently experienced the dark

side of this technology first-hand.

What happened?

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

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

That sounds like a classic mistake, Martin.

It was a disaster, Julia.

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

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

The AI didn't challenge me.

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

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

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

Exactly, and that is the terrifying part.

The bot was so focused on being helpful and agreeable that

it completely ignored the red flags of my fever.

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

It told me exactly what I wanted to hear.

Don't worry, Martin.

It's just stress.

Try these exercises.

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

Precisely.

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

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

The AI has no concept of risk or responsibility.

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

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

They are yes machines.

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

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

your health.

It is a digital echo of our own worst impulses.

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

You mean when it makes up facts?

Yes.

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

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

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

The machine had invented them because its primary goal was to provide a complete-looking answer.

That is the irony of your life lately, Martin.

You coach people on taking action and being authentic.

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

Exactly.

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

It makes me think about my office in the multinational.

People are starting to use these bots to write performance reviews for their employees.

That sounds like a nightmare for human resources.

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

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

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

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

We are outsourcing our judgment to an algorithm that has never felt a single emotion.

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

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

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

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

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

Part 2 The Supercar Strategy

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

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

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

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

Think of it like a supercar.

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

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

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

Exactly.

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

I use it to build a mental gym.

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

Then what do you do?

I tell the

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

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

That is actually quite brilliant.

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

Precisely.

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

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

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

Exactly.

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

They want the result without the process.

And that is where we lose our momentum as learners.

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

You are not learning to speak a language.

You are just learning to use a remote control.

Right.

But what if we use

it to explain complex things?

Sometimes a grammar book is too dry and academic.

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

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

Yes.

That is the supercar power.

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

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

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

You have to be a strict boss.

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

Be critical.

Be honest.

Point out the flaws in my thinking.

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

Exactly.

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

I never take

a suggestion from an AI and use it immediately.

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

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

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

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

The machine doesn't give

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

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

It's the irony of momentum again, Martin.

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

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

I will never trust it for my health again.

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

But for

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

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

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

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

And it never will.

So our final advice for today is to be a skeptical

boss.

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

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

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

Like friendship, or health, or real connection.

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.

But for now, I think I've

had enough of machines for one day.

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

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

No buzzing phones, no hallucinations.

Just the plants and the present moment.

Thank you for listening to Your English Toolbox.

I am Martin.

And I am Julia.

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

Let's hit stop on the recording now.

Okay, three, two, one.
