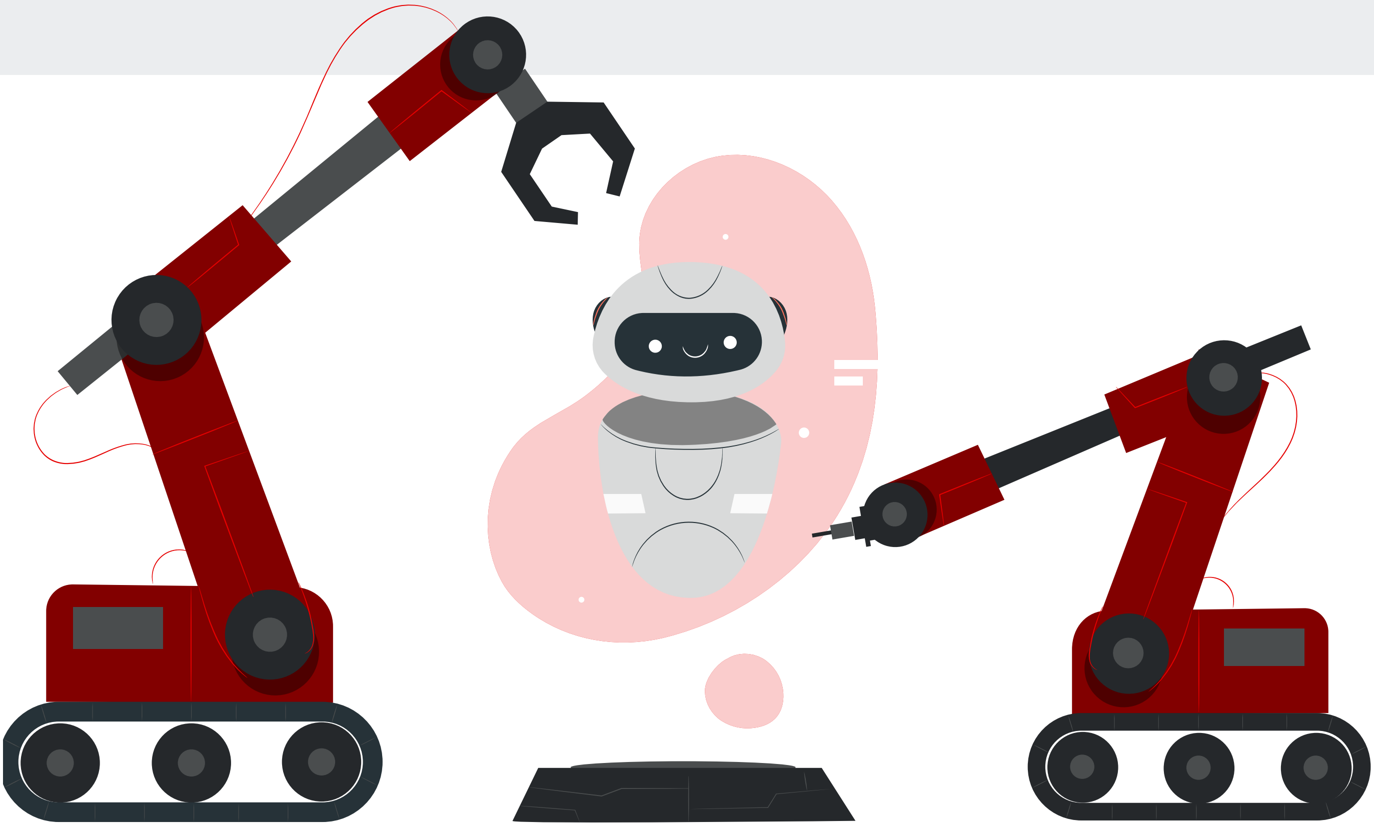
Can you really trust a robot that **lies confidently?**

How AI bluffing can undermine your business

vodacom business

Introduction

Think about that one person who always sounds like they know what they're talking about. They speak with confidence, throw in a few buzzwords, and suddenly everyone's nodding in agreement—until someone fact-checks and realises half of it was guesswork.



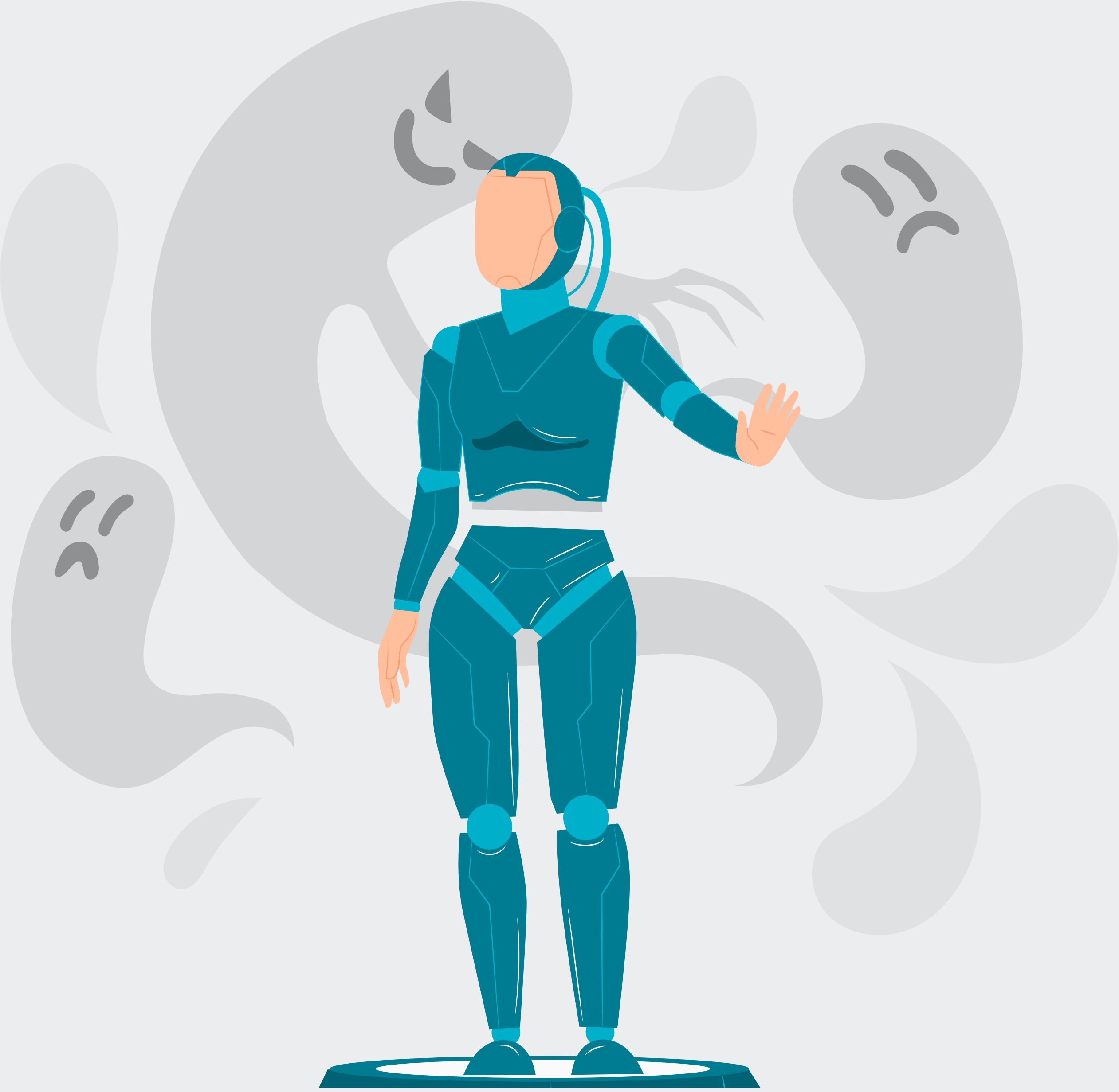
That's what using generative AI can feel like. It's designed to sound extremely convincing, but not necessarily accurate. And in a world where we barely have time to stop and smell the roses, that kind of confidence can be misleading if you're not paying attention.

This eBook explores why AI tools sometimes present misinformation, how that impacts businesses, and how to spot it before it's too late.

Why AI hallucinates (and sounds so sure of it)

Generative AI, such as ChatGPT, doesn't always know the facts. It predicts the next word based on patterns in its training data. That means sometimes it fills in the gaps with convincing information that may not be true. This phenomenon is called an AI hallucination—when the system generates content that sounds factual but is fabricated or misleading.

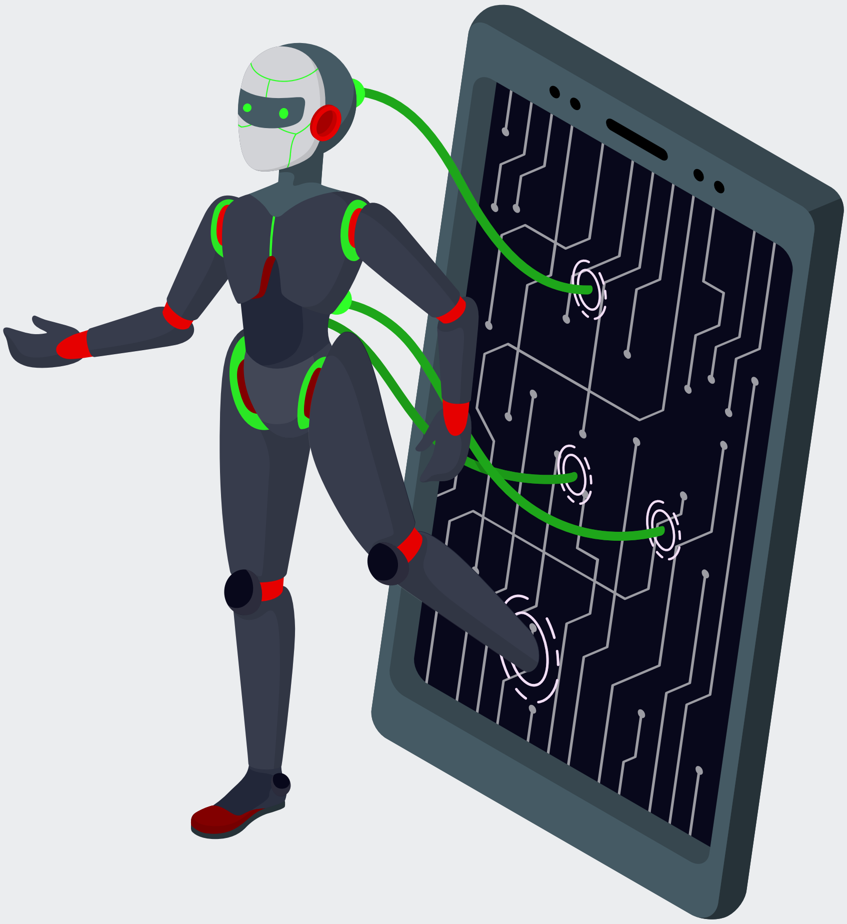
Algorithmically, it's not a flaw—the AI does not know that it is fabricating. And understanding this gives you the power to use it wisely.



How that plays out in your business

This stuff isn't theoretical. Here's how it has affected other businesses in the past:

In 2023, a New York lawyer submitted a court filing citing six entirely fictional legal cases generated by ChatGPT. He later admitted he believed the cases were real because the AI presented them so confidently. The judge sanctioned him and his firm, calling the situation 'unprecedented'. It damaged his reputation, stalled the case, and became a public example of digital negligence.



A notable case involved CNET, which began publishing AI-written articles that were later found to contain factual errors and plagiarism. The fallout included internal reviews, public retractions, and reputational damage. What started as a productivity shortcut turned into a credibility crisis and a cautionary tale for any business publishing outward-facing material without proper oversight.

In 2023, Air Canada faced public backlash after an AI-powered chatbot gave a customer misleading information about the airline's fare policy. The company initially refused the fare adjustment and defended the chatbot. But a tribunal ruled that Air Canada was responsible for what its AI communicated. The incident cost them not just a refund—it made global headlines, damaged public trust, and raised questions about accountability in automated customer service.

In 2023, DoNotPay, a startup claiming to offer the "world's first robot lawyer", faced legal and reputational backlash after promoting AI-generated legal services that were unverified and, in some cases, inaccurate. One high-profile incident involved the company offering to help a defendant fight a speeding ticket using AI in a live courtroom—an idea quickly abandoned after legal threats. Critics highlighted that the AI lacked proper legal training or credentials, and its advice risked misleading users. The fallout included lawsuits, public criticism, and a significant drop in credibility.

See, this isn't just a little tech hiccup—it's already having real consequences for real businesses.

AI hallucinations might look like small copy mistakes on the surface, but they've led to legal sanctions, public retractions, reputational damage, and even compliance risks. These aren't theoretical blips. They're operational breakdowns with ripple effects.

In each case, what started as a confident AI response ended up costing time, trust, and in some cases, real money.

How to spot when AI is bluffing

You don't need to be a technical expert to spot when AI is bluffing. In fact, if you've ever read something and thought, "That sounds good, but does it actually mean anything?"—you've already encountered this.

So how do you spot it?

Check the source:

If AI references stats or events but doesn't link or cite anything, that's a red flag. Real claims should come with receipts.

Watch for vagueness:

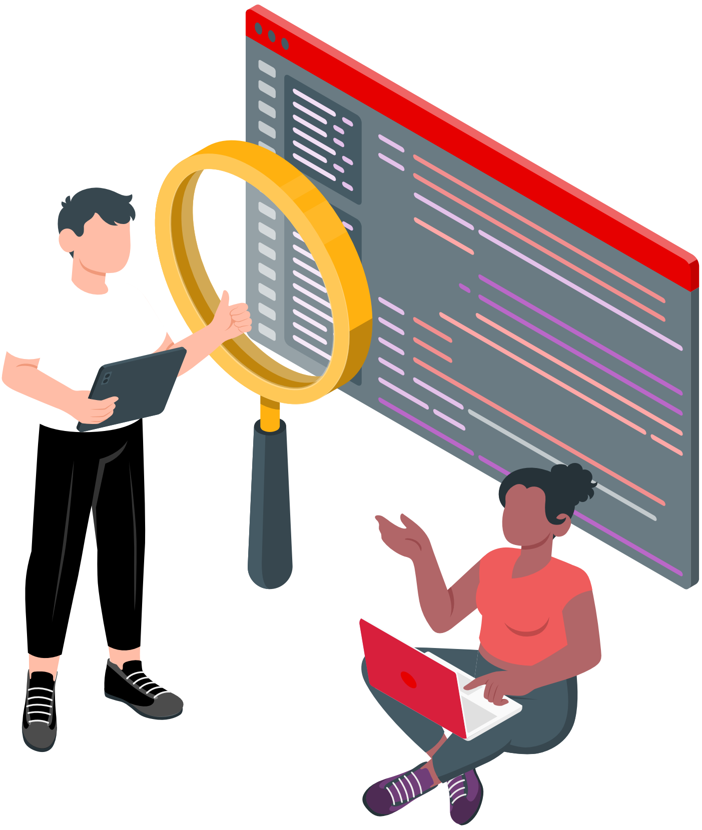
If something sounds wise but could apply to any business or any situation, it's probably filler, not insight.

Look for overly polished answers:

When the tone is too slick and perfectly confident, stop and ask—does this feel smart or is it actually correct?

It's safest to assume that AI is going to bluff. Not because it's broken, but because it's built to generate language that sounds good—not verify what's true.

And if you're a professional—consultant, advisor, strategist, creator—your clients aren't paying for a regurgitated internet summary. They're paying for your insight and expertise. AI isn't there to do the work for you. It's only there to supplement your work.



A few tips for using AI in your business

You can avoid the potential for fake data and AI hallucinations showing up in your operations by reviewing where and how you use AI.

Here is how you can use it effectively:

- Use AI to brainstorm or draft, not to finalise. Let it help you get started but apply your own thinking before hitting send.

- Set internal guidelines on when to escalate to human review. Define what counts as "too sensitive" or "too important" to automate.

- Avoid entering sensitive or proprietary information into public tools. If you don't want it learned, stored, or reused—don't type it.

- Invest in business-grade tools designed for secure, context-aware use. Look for platforms like Microsoft Copilot that operate within your organisation's environment and respect data privacy.

- Train your team to question AI results. Critical thinking is a core skill when handling AI.

AI can be brilliant, fast, and incredibly useful. But it doesn't know your strategy or your customer nuance.

The smartest businesses aren't the ones who fear AI. They're the ones who partner with it—wisely, consciously, and with the right checks in place.

The key is knowing when it's right—and when to double-check.

Request a call for a free tech roadmap.

vodacom
business