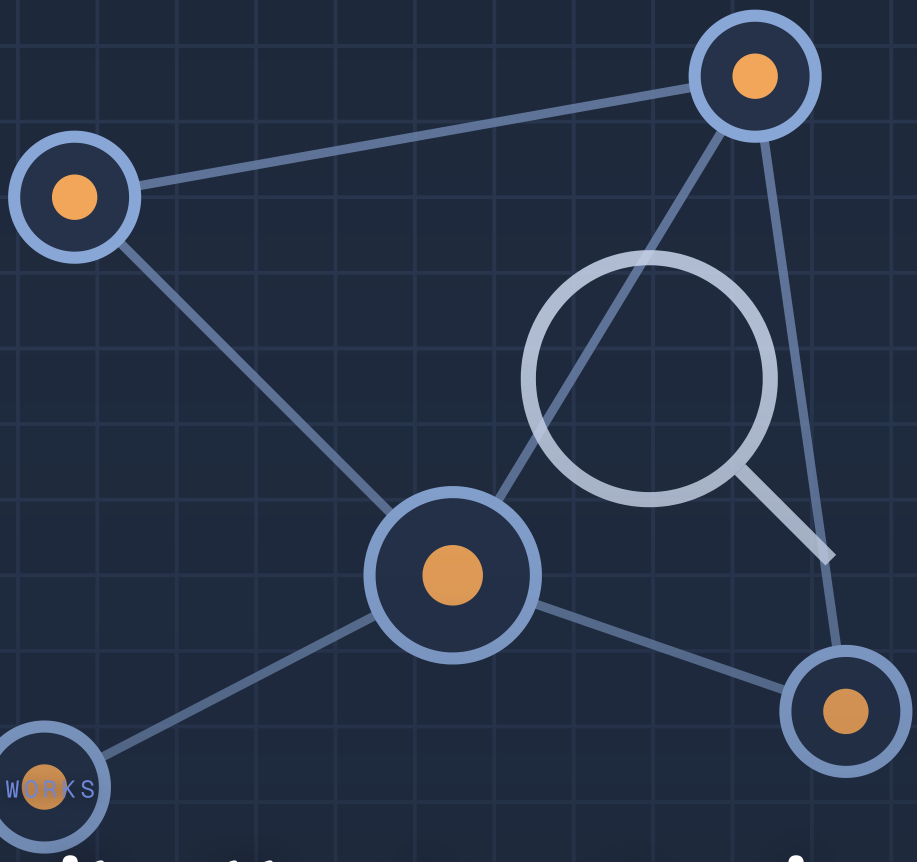
Vocabotics

A VOCABOTICS EXPLAINER



HOW AI REALLY WORKS

What 'hallucination' really is

AI can state something completely false with total confidence. Here's what that really means, in plain English, and where it bites hardest — money, health, and the law.

For anyone who wants to understand

Shipped — measured, gate-clean

COVER: A CLEAN NODE-AND-EDGE DIAGRAM

Vocabotics

A VOCABOTICS EXPLAINER

What 'hallucination' really is

AI can state something completely false with total confidence. Here's what that really means, in plain English, and where it bites hardest — money, health, and the law.

PREPARED

19 June 2026

VERIFIED BY

Jon Ross — Founder, Vocabotics — 15 years building safety-critical AI

HOW TO READ IT

Drafted with AI, verified by a named human. Sources at the back.

01 The short version

"Hallucination" is the polite industry word for a simple, important fact: an AI can tell you something that is completely made up, in fluent, confident, professional-sounding language. It isn't lying — it has no idea it's wrong. And that's exactly why it matters most in the moments you can least afford it.

The one sentence to remember

AI does not know things the way a person does. It predicts likely-sounding text. Usually the likely-sounding answer is also true — but not always, and it sounds exactly the same either way.

02 Why it happens (the honest, jargon-free version)

A chatbot is, underneath, a very sophisticated pattern-completer. It has read an enormous amount of text and learned what words tend to follow other words. When you ask it something, it produces the most plausible continuation.

That's why it's so good at writing and explaining — plausible and correct usually line up. But when it *doesn't* know something, it doesn't stop and say so. It produces a plausible answer anyway, because plausible text is the only thing it makes. A made-up court case, a fake statistic, a citation to a book that doesn't exist — all delivered with the same calm confidence as the true bits.

> ***Where an AI system's output is used to make decisions about people, its accuracy is not optional — it carries real legal and ethical weight.***
>
> Guidance on AI and data protection · UK Information Commissioner's Office · as of June 2026

What it feels like

- It looks the answer up
- It knows what is true
- Confidence means correctness

VS

What actually happens

- It predicts likely next words
- It has no fact-store to check against
- Confident and wrong read the same

Why hallucination happens: a language model predicts plausible next words — it does not look an answer up in a store of verified facts. So a confident sentence and a correct sentence can read exactly the same.

03 Where it bites hardest

The danger isn't the mistake itself — it's the *confidence*. You can't hear uncertainty in the answer, so you're most exposed exactly where you're least able to check:

Health. A confident, wrong explanation of a symptom or a medication.

Money and tax. A made-up figure, allowance, or rule stated as fact.

The law. Invented cases and citations — this has already caught out professionals who trusted an answer without checking.

Anything you'll act on but can't easily verify. The unfamiliar is where hallucination does the most damage.

04 How to protect yourself — in ten seconds

You don't need to become an expert. You need one habit: **ask it to show its work, then check the source yourself.**

"What's your source for that? Give me a link I can check."

"How confident are you, and what would make this wrong?"

For anything important, confirm the answer against a trustworthy human or an official source before you rely on it.

If it can't give you a checkable source, treat the claim as unconfirmed — not as false necessarily, but not as fact either.

Ask for the source

then actually click it

Ask twice

watch for a different 'fact' the second time

Verify before you act

figures, names, dates, law

The ten-second defence — three habits that catch almost every confident invention.

05 Why this shapes how we build

We think an AI that *can't tell you when it doesn't know* is dangerous in exactly the settings that matter. So the systems we build are designed to **cite their source or refuse to answer** — no confident guessing where being wrong is expensive. That's the honest standard we hold our own work to, and the one worth holding any AI to.

The takeaway is calm, not scary: AI is a superb assistant and a poor oracle. Let it draft and explain; keep a human — and a source — between its confidence and your decision.

06 Sources, glossary & the ledger

Where every claim comes from — and what the plain words mean.

Sources

01 **Guidance on AI, accuracy and statistical accuracy** — UK Information Commissioner's Office (as of June 2026)
https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/

Vocabotics

THE HONESTY LEDGER

🟢 Shipped — measured, gate-clean.

🔬 Research — measured, partial; the gap is shown.

🔭 Vision — designed, honest; not yet built.

The bigger claim goes up a tier, never down. We show the eval beside the demo — the tiering is the credibility.

See which AI tools to actually trust, and for what

Now you know why AI can be wrong, here's our honest, no-affiliate view of which tools are worth using — and where to keep a human in the loop.

/resources/ai-for-everyone/which-ai-should-you-actually-use-2026

Nothing faked.

Drafted with AI and verified by a named human. A clear explanation you can act on — the mechanism in plain English, the limits marked.

Vocabotics · calm, clear, candid · no client names · global rail & metro heritage