A FYTAHQ COMPANION RESOURCE

AI From Scratch

A Plain-Language Course
for people who did not grow up writing software

LESSON 3 OF 8

"Three buckets: what it does well, what it does badly, and what it does badly while sounding completely confident."

What AI Gets Wrong

ISSUE · JUNE 2026 · FACELESSYOUTUBEAUTOMATIONHQ.COM

Lesson 3 — What AI Gets Wrong

WHAT YOU WILL KNOW AFTER THIS LESSON

- The three buckets every AI task falls into — does well, does badly, and the dangerous middle
- Why that third bucket, confident and wrong at the same time, is the one to watch
- What "hallucination" actually means, in plain words, with one real example
- The one strength of AI almost nobody mentions — it never gets tired of you
- How much trust to place in an answer, and when to slow down and check it yourself

You do not have to memorise the lists below. The three buckets are the whole takeaway. Once the buckets are in your head, the rest of this lesson is just a checklist you already understand.

PLAIN-LANGUAGE EXPLANATION

AI comes in three buckets, not one yes-or-no answer

People ask "Is AI any good?" as if there were a single answer. There is not. The honest reply depends entirely on what you ask it to do. The very same tool that writes you a warm, well-judged email in four seconds will also invent a fact in the next sentence and hand it to you with total confidence.

So it helps to stop sorting AI into "good" or "bad", and instead sort each **task** into one of three buckets. The first bucket holds the things it genuinely does well. The second holds the things it does badly in a way you spot straight away. The third — the one worth real attention — holds the things it does badly while sounding completely sure of itself.

Here are the three buckets at a glance. The rest of the lesson simply walks through each one in turn.

Does well

- Writing first drafts
- Summarising long text
- Rough translation
- Brainstorming ideas
- Explaining concepts

Use it freely. Skim, tidy, send.

Does badly (you notice)

- Exact arithmetic
- Very recent events
- Small local facts
- Predicting the future
- Strict formatting

It is wrong, and it shows. Low risk.

Confident, but wrong

- Medical advice
- Legal advice
- Money and tax advice
- Invented sources
- Specific statistics

Sounds right, may not be. Always check.

WHY THIS MATTERS

Once you can glance at a task and quietly drop it into the right bucket, you stop asking the unanswerable question — "can I trust AI?" — and start asking the useful one: "can I trust it *for this*?" That single shift is what separates the people who get burned by AI from the people who calmly get good value out of it every day.

Lesson 3 — What AI Gets Wrong

Bucket one — the things it genuinely does well

This is the bucket that earned AI its reputation, and the reputation is deserved. When the job is to take words and turn them into more words, today's tools are quick, tireless, and good enough that you only need to tidy the result rather than start from scratch.

- **Writing first drafts** of almost anything — emails, posts, letters, reports, even a speech for a retirement party.
- **Summarising** long documents, articles, or the transcript of a video into a few clear sentences.
- **Translating** between major languages — rough around the edges, but usually good enough to understand the meaning.
- **Brainstorming** lists of ideas, names, alternatives, or ways to structure something you are stuck on.
- **Explaining a concept** at whatever level you ask for — "as if I were ten" or "as if I had a degree in it".
- **Reformatting** text you already have — turning bullet points into paragraphs, or a wall of text into a tidy list.

Notice what these share. There is no single correct answer, the stakes are low, and you can see the result with your own eyes and judge it. That is the sweet spot.

> **A plain example.** You paste in a rambling three-page email from a colleague and ask, "What is the one thing they actually want me to do?" A couple of seconds later you have a clear, one-line answer. There is no single correct wording, the cost of a small slip is nothing, and you can check it against the email sitting right in front of you. That is bucket one working exactly as it should.

Bucket two — the things it does badly, where you notice right away

The second bucket is almost reassuring. Here the AI gets things wrong, but it gets them wrong in plain sight, so the damage is small. You spot the mistake, roll your eyes, and move on.

- **Maths** — yes, really, and yes, it is ironic for a computer. Basic arithmetic gets shaky with large numbers, especially when the sum is dressed up as a word problem.
- **Recent events** — the model's knowledge stops at a fixed date (its "training cutoff", explained on the next page), so anything after that is simply unknown to it.
- **Very local facts** — the name of a small shop in your town, or last week's local news, is usually beyond it.
- **Predictions** — it does not know the future. It only knows how to make a guess sound plausible, which is not the same thing at all.
- **Strict formatting** — asked to hold to an exact word count or a rigid template, it tends to drift, add a stray line, or quietly ignore the rule you set.

Because these errors are obvious, this bucket needs little discipline from you. The one that does is on the next page.

Bucket three — wrong, while sounding completely confident

This is the bucket that matters. The trouble is not that the AI is wrong — bucket two is wrong too. The trouble is that here it is wrong while sounding calm, fluent, and certain. There is no nervous pause, no "I think", no hedge. It states the false thing in exactly the same confident voice it uses for the true things.

These are the topics where a confident-but-wrong answer can actually cost you something:

— **Medical advice** — paragraphs that read like good medicine, and sometimes are not.
— **Legal advice** — the same problem, and usually more expensive when it is wrong.
— **Financial and tax advice** — which depends on your country, age, and situation, none of which the AI actually knows.
— **Invented sources** — citations and references that look entirely real but point to nothing that exists.
— **Invented quotes** — real-sounding words placed in the mouth of a real person who never said them.
— **Specific statistics** — confident percentages, counts, and dates that were never looked up anywhere.

When an AI makes something up like this and presents it as fact, there is a name for it. It is called a hallucination, and it is the single most important idea in this lesson.

Here is the part that catches people out: the AI is not doing this to deceive you. It has no idea it is wrong. It was built to produce text that *sounds* like a good answer, and a made-up citation sounds every bit as plausible as a real one. So the quiet job of checking simply moves to you.

PLAIN-LANGUAGE KEY — THE TERMS USED IN THIS LESSON

Prompt
The thing you type in — your question, request, or instruction to the AI. "Prompt" is just a fancier word for "what you asked".

Output (also "response")
The answer the AI hands back. It can be a sentence, a page, a list, or an image — whatever you asked for.

Hallucination
When the AI makes something up and states it confidently as fact — a fake source, a wrong number, a quote nobody said. Not lying on purpose; it simply cannot tell the difference.

Training cutoff (knowledge cutoff)
The date after which the model has learned nothing. Ask about last week and it will guess, because its reading of the world stopped months ago.

Citation
A reference to a source — a study, a book, a web page. AI is alarmingly good at inventing ones that look perfectly genuine.

Fact-check (verify)
To look something up yourself before trusting it — open one real source and confirm the claim with your own eyes. The single habit that quietly defuses almost every hallucination.

A CONCRETE EXAMPLE — WHAT A HALLUCINATION LOOKS LIKE

Imagine you ask an AI: "What were the top three reasons the Roman Empire fell?" It gives you a tidy, well-written answer — economic decline, military overstretch, political instability. So far, so good. This is bucket one, and it is genuinely useful.

Now you ask a follow-up: "Cite three peer-reviewed papers from the last five years that support this." The AI replies with three papers — real-looking author names, real-looking journals, real-looking years. Every detail looks correct.

None of those three papers exist. The AI did not look anything up. It simply produced text that looks exactly like what a real citation looks like, because that is the only thing it was ever built to do.

You will not notice unless you check. Many people never check. That gap — between an answer that sounds right and an answer that is right — is the entire risk this lesson is about.

The fix is simple and takes a minute. Ask for the sources, then look up just one of them. If the first source does not exist, it is safe to assume the others do not either, and the whole answer needs a second opinion.

THE STRENGTH ALMOST NOBODY MENTIONS

Infinite patience. The AI will explain the same thing to you fifteen times, in fifteen different ways, and never once sigh, glance at the clock, or make you feel slow.

For anyone who feels awkward asking a real person the same question twice, this is genuinely new. You can ask "and what does that word mean?" as many times as you like. Nobody is keeping score, and nobody is judging you.

THREE THINGS TO LOOK OUT FOR

1. **Confidence is not correctness.** AI writing is smooth, and smooth writing quietly feels true, which slips past your usual scepticism. Train yourself the other way: when the answer sounds its most polished and sure, that is precisely the moment to slow down and verify.

2. **Numbers are a flag.** Any time the AI gives you a specific figure — a percentage, a date, an amount, a population — treat it as a prompt to check at least that one. The AI has no database of facts; it predicts what a likely-sounding number would be.

3. **Recent means risky.** If the topic is anything from the last several months, the model may never have read about it, yet it will still answer — sometimes with confident fiction. For fresh events, an ordinary search engine is still the more reliable tool.

WHAT IS SAFE TO SKIP

You do not need to know any of the following to use AI sensibly and read its answers with the right amount of trust:

- The technical reason hallucinations happen — it involves probability, and it changes nothing about what you do
- Whether a bigger model is better — sometimes yes, often no, and never worth losing sleep over
- Benchmark scores with names like MMLU or GSM-8K — fascinating to engineers, irrelevant to you
- The difference between "reasoning models" and "chat models" — mostly marketing decoration as of 2026
- Which provider is ahead this month — it changes monthly, and the three buckets stay the same regardless

You are allowed to use AI for the things it does well, and skip it for the things it does badly. The list of buckets is a practical guide, not a moral hierarchy, and nobody is grading how you use it.

NEXT TUESDAY

Lesson 4 — Why Some Are Excited, And Why Some Are Worried

A balanced, honest look at the AI debate — the real reasons people are excited and the real reasons people are worried, laid side by side. No cheerleading, no doom. Free on the resources page, with no email and no sign-up required.

PREFER A LIVE CONVERSATION?

One-on-one sessions with a real person

If you would rather have a real human walk you through this material in a live one-on-one session than read on your own, the mrocon side of our world offers exactly that — the same philosophy as this course, simply guided. One, two, or three hours, depending on how deep you want to go.

`mrocon.at/pricing →`

COMPANION VIDEO

Lesson 3 — Video Version on the channel

The Lesson 3 video is live on YouTube today — the same material, in video form. The full "AI From Scratch" playlist updates every Tuesday for the rest of the series. Watch what fits your evening, and read what fits your morning.

`facelessyoutubeautomationhq.com →`

Sources + version note: *Content valid as of June 2026. Reference categories: ChatGPT / OpenAI · Claude / Anthropic · Gemini / Google · documented hallucination behaviour across major chat assistants · EU AI Act (transparency obligations). AI capabilities shift every few months — newer models handle some of the "does badly" items better — so this lesson stays at the level of categories and principles, which remain stable between update cycles.*