

The AI Bootcamp

18

6 Questions to Ask Any Long Document

In this chapter you'll learn

Where long documents actually fail, 6 questions to put to any of them, and a check that takes seconds

The AI Bootcamp

THE PROBLEM

Most AI education is a pile of disconnected tips. A LinkedIn carousel here, a YouTube tutorial there, a \$2,000 cohort course that goes out of date before the refund window closes. None of it is built to be read in order, so nothing compounds. You finish twelve videos and can recite twelve facts, not perform one new skill.

THE FIX

The AI Bootcamp is built the other way. 90 skills, packed 2 to 3 per chapter, 36 chapters, 3 tracks. Every chapter states what it needs from you and what it hands to the chapter after it, so skill 47 only works because skill 12 already landed. Read it in order and by the end you are not collecting facts about AI, you are running work through it.

LETTER FROM THE EDITOR

We built this because we kept meeting smart people who felt behind on AI for no reason other than nobody had organized the material for them.

This book does not assume you are technical, and it does not water anything down either. Every chapter earns its place: a real skill, a real exercise, a one-page companion you can flip back to in month three when you have forgotten the details but still need the answer.

If a chapter loses you, or a chapter changes how you work, I want to hear about it either way.

Kris

Editor, AI Central

Write to me at editor@thecentral.ai

THE 3 TRACKS

AI 101

Get running

Pick a tool, set it up, write a clear instruction, use it safely at work.

AI 202

Prompt properly

The framework behind every good instruction, and catching AI when it is wrong.

AI 303

Build without watching it

Connect AI to your tools, and automations that run while you are not looking.

IN THIS LESSON YOU'LL LEARN

By the end of this chapter you will know which part of a long document an assistant is most likely to read badly, you will have 6 questions you can put to any document instead of asking for a summary, and you will have a check that tells you in seconds whether a claim came out of the document or out of the model.

Ask for a summary of a 90-page document and you get back something fluent, confident and well organised, which has quietly left out the one paragraph that would have changed your decision. You cannot tell which paragraph, because the summary is now the only version you have read. Nobody has ever been caught by a bad summary. People get caught by a good one.

CONTENTS	18.1	Reading the whole thing and remembering all of it are different claims
	18.2	A summary is the least valuable thing a long document can give you
	18.3	6 questions to put to any long document
	18.4	Make it quote, then go and find the quote

TOPICS REQUIRED

Chapter 17. You should have a prompt library started, even if it holds only 2 templates, and the framework from Chapter 13 in your head. Bring one long document from your own work: a contract, a board pack, a policy, a research report, something you would rather not read twice.

AFTER THIS CHAPTER

Chapter 19 assumes you can put a long document in front of your assistant and question it rather than compress it. It goes after the part this chapter deliberately leaves alone: the numbers inside the document.

18.1 Reading the whole thing and remembering all of it are different claims

Every one of these tools will accept a document far longer than you would ever read yourself. None of them promises even attention across all of it, and the research keeps confirming that you do not get it.

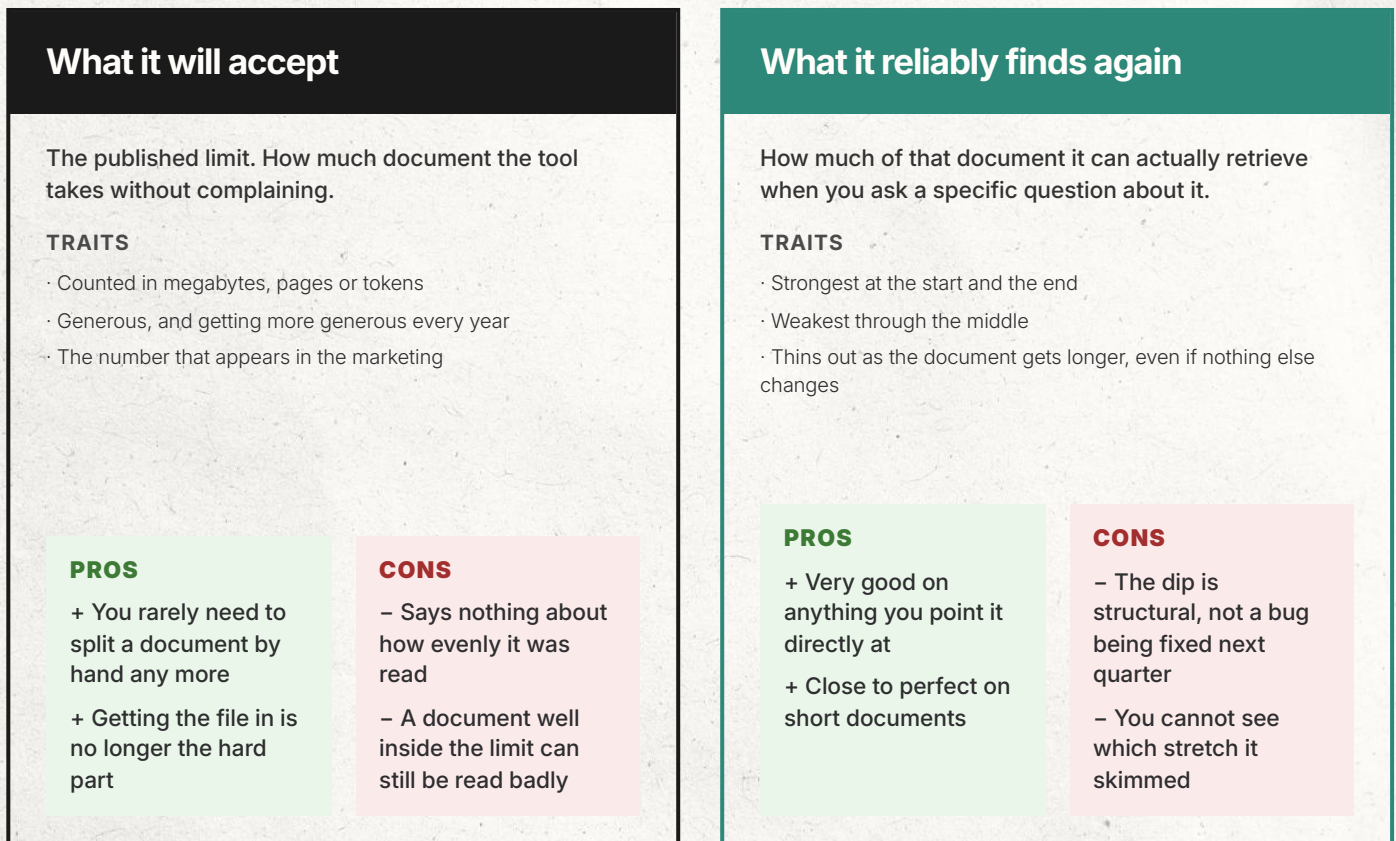


Fig 18.1 Capacity is the number the vendor publishes. Recall is what you actually get back when you ask. The gap between them is where a long document fails you without saying so.

18.1

Claude. Up to 500MB per file and 20 files per chat, PDFs to 1,000 pages. Past 100 pages it stops reading the charts and images and reads only the text.

ChatGPT. A hard 512MB per file, plus a separate cap of roughly 2 million tokens of text, so a small file can still be too long.

Gemini. 10 files per prompt, 100MB each. Google describes its 1 million token window as around 1,500 pages of text.

Copilot. In Word it summarises documents up to roughly 80,000 words, and Microsoft states that it does not always cite the later parts of one.

Here is the problem this chapter exists to fix. You hand over a 90-page document, you get back four confident paragraphs, and you have no idea which parts of those 90 pages the assistant actually looked at.

Capacity is genuinely no longer the constraint. All four assistants now take documents longer than most people will ever hand them, and the specifics sit in the margin of this page. Getting the file in was Chapter 7's job and it is the easy half.

Recall is a different quantity, and it behaves in a way that has been documented since 2023 and has survived every generation of model since. Accuracy is highest for material at the very start of a long input and at the very end, and lowest through the middle. One 2026 evaluation ran 17 long-context models and found every one of them losing accuracy as the input grew. Larger context windows did not remove the effect. They gave you more middle.

A 2026 study went further and isolated length itself. Even when the surrounding filler was stripped down to blank space, so there was nothing to distract or confuse, performance still fell as the input got longer. Length is not free. A document of 200 pages is harder for the tool to work with than the same content in 20, quite apart from what is in it.

The assistants are not identical here. Gemini in particular now handles single-fact retrieval from very long inputs well wherever the fact sits, and every model on the market is better at this than its predecessor. What no production model has done is remove position effects, and the gap between finding one fact and reasoning over several scattered ones is where the remaining failures live.

The practical consequence is small and worth building a habit around. Before you ask anything, know roughly where in the document the thing you care about sits. If it sits in the middle of a long file, that is the part you check first, not last.

18.2

A summary is the least valuable thing a long document can give you

Summarise this is the first thing almost everyone types and close to the last thing an analyst would. It is also the only request whose output you have no way of checking.

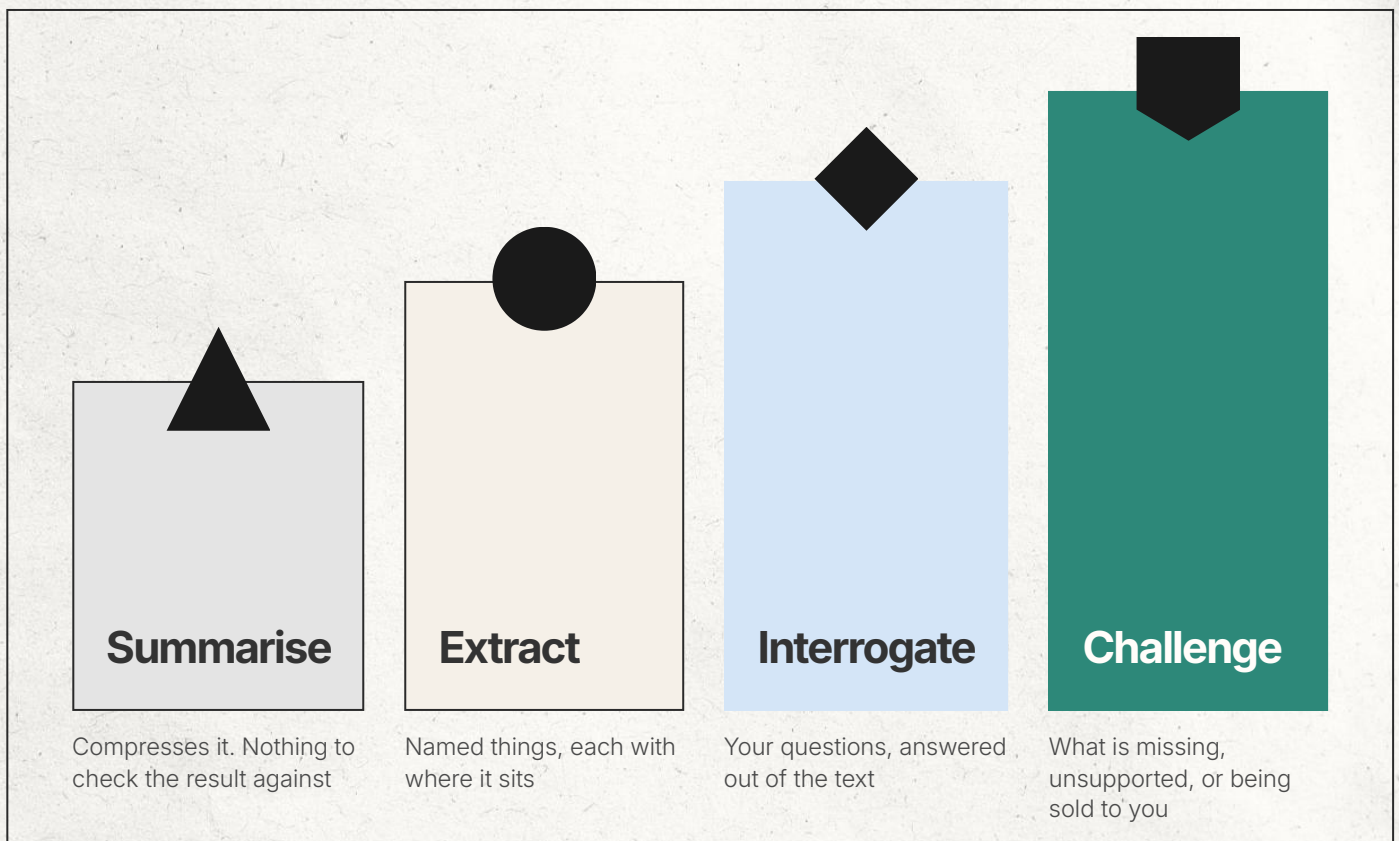


Fig 18.2 Four altitudes of asking. The bar heights are a judgement about how much each one gets you, not a measurement of anything, and only the bottom three produce something you can go and verify.

18.2

The summary trap.

A summary is judged on how it reads. It is never judged on what it left out, because you cannot see that.

Extraction is checkable. Ask for named things and where they sit, and you can go and look.

Lossy on purpose. You asked it to throw most of the document away. It did exactly that, using its judgement about what mattered, not yours.

A summary fails silently. That is the whole objection to it. A summary that dropped the indemnity clause, the covenant, the one dissenting paragraph in the appendix, reads exactly like a summary that kept them, and you will never know which one you are holding.

It is also the request that plays worst to the tool. Research published this year found that faithfulness in long generated summaries falls off toward the end of the output, so the later paragraphs of a long summary are the least trustworthy part of it, which is not where anyone expects the weak spot to be. And omission has been studied far less than invention, even though at work it is the more expensive failure. An invented fact can be spotted. An absent one cannot.

Extraction is a better question because it produces something falsifiable. Every obligation with a date attached. Every named party. Every commitment the document makes that has a number or a deadline in it. Ask for those as a list with locations and you have output that can be wrong in a way you can detect, which is the only kind of output worth having.

Interrogation is better still, and it is what actually separates an analyst from a reader. An analyst does not compress a document. They arrive with a question, put it to the document, and use the document to answer it. The 90 pages are not the deliverable. Whatever you needed to know is.

Challenge is the level almost nobody reaches. What is this document not telling me. Which of these claims does it actually support and which does it simply assert. What would somebody who disagreed with this say. You can ask an assistant all three, and it will answer them, and it is a slightly strange fact about this technology that almost nobody does.

For the next few pages, stop typing summarise. If you want a way in, use the 6 questions on the facing pages, in order.

18.3

6 questions to put to any long document

Not a summary request. A reading pass you run in the same order every time, on every document, which makes it precisely the sort of thing Chapter 17 told you to save as a template.

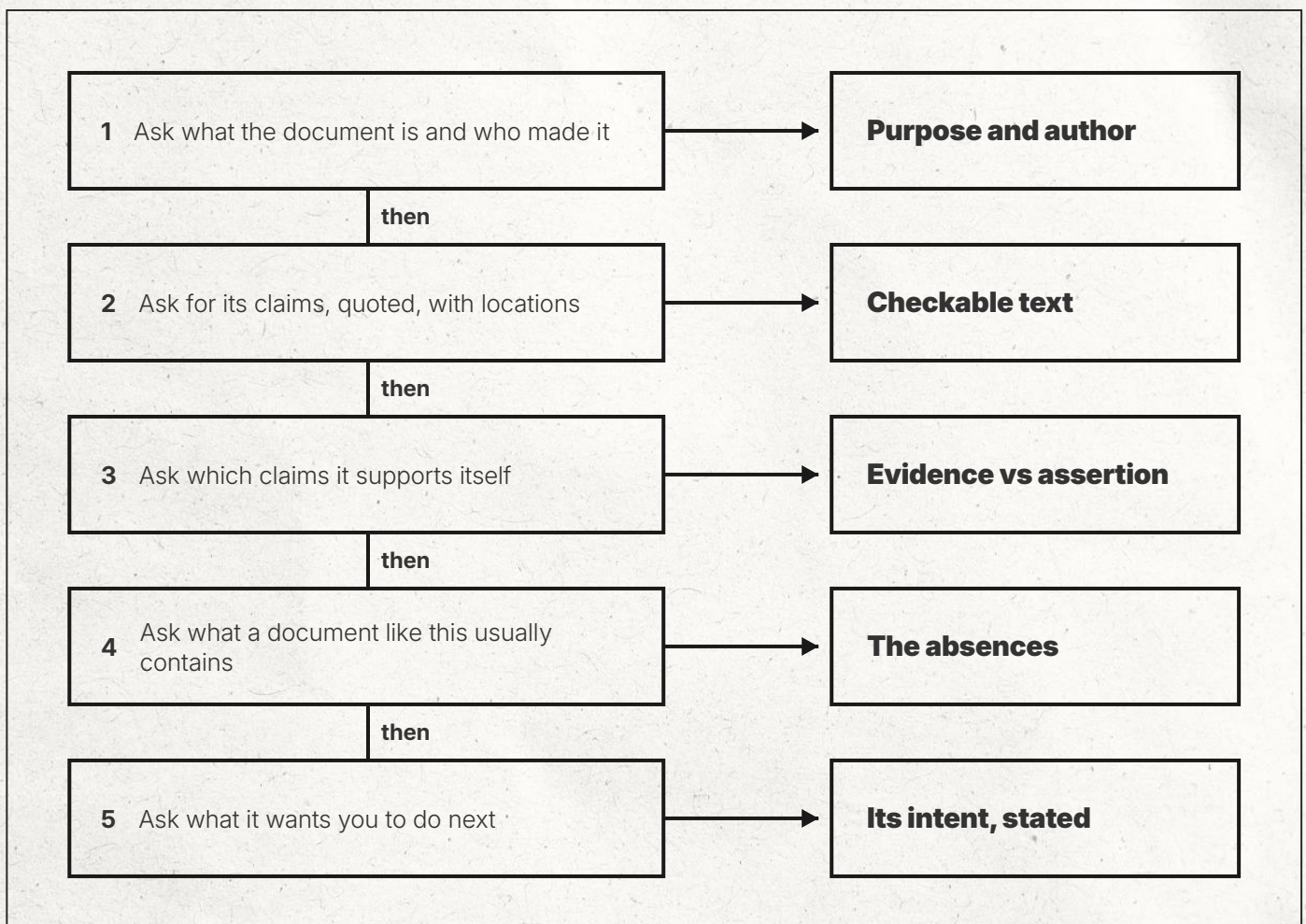


Fig 18.3 The pass, in order. Step 2 is the one people skip, and it is the only step that makes everything after it checkable.

18.3

Ask about absence. What a document leaves out is very often the entire point of the document.

Ask about intent. Almost nothing over 10 pages is neutral. Something in there is arguing for something.

One question per message.
Six questions in one prompt buys you six shallow answers. Send them one at a time.

Question 1. What is this document, who produced it, and what does it want its reader to do. Ask this first even when you think you know, because the answer frames every other answer, and an assistant will often name an audience or a purpose you had not considered.

Question 2. What are the claims this document most wants me to accept. Quote each one exactly and tell me where in the document it sits. This is the load-bearing question of the whole pass. You are asking for words that exist rather than words about words, and section 18.4 is entirely about what you then do with them.

Question 3. Which of those claims does the document support with evidence inside itself, and which does it simply assert. This one line does most of the analytical work. Documents that are trying to persuade you tend to be evidence-heavy exactly where it does not matter and assertion-only exactly where it does.

Question 4. What would a document of this kind normally contain that this one does not. Then question 5. Where does it contradict itself, or quietly change what a term means partway through. These two catch the things a summary is structurally incapable of catching, because both are about what is not on the page.

Question 6. If I act on only one thing in here, what should it be, and what would have to be true for that to be the wrong call. It gives you the decision and its own failure condition in one answer, which is a more useful pair than a conclusion on its own.

Run these as 6 messages in one thread, not one message with 6 parts. And on anything genuinely long, run the pass over the document in sections rather than in one go. Chapter 16 made the case for splitting a big job into steps, and a 200-page document is the clearest example of it there is.

18.4 Make it quote, then go and find the quote

One habit separates a finding you can act on from a sentence that merely sounds like one. Ask for the actual words, then search the actual document for them.

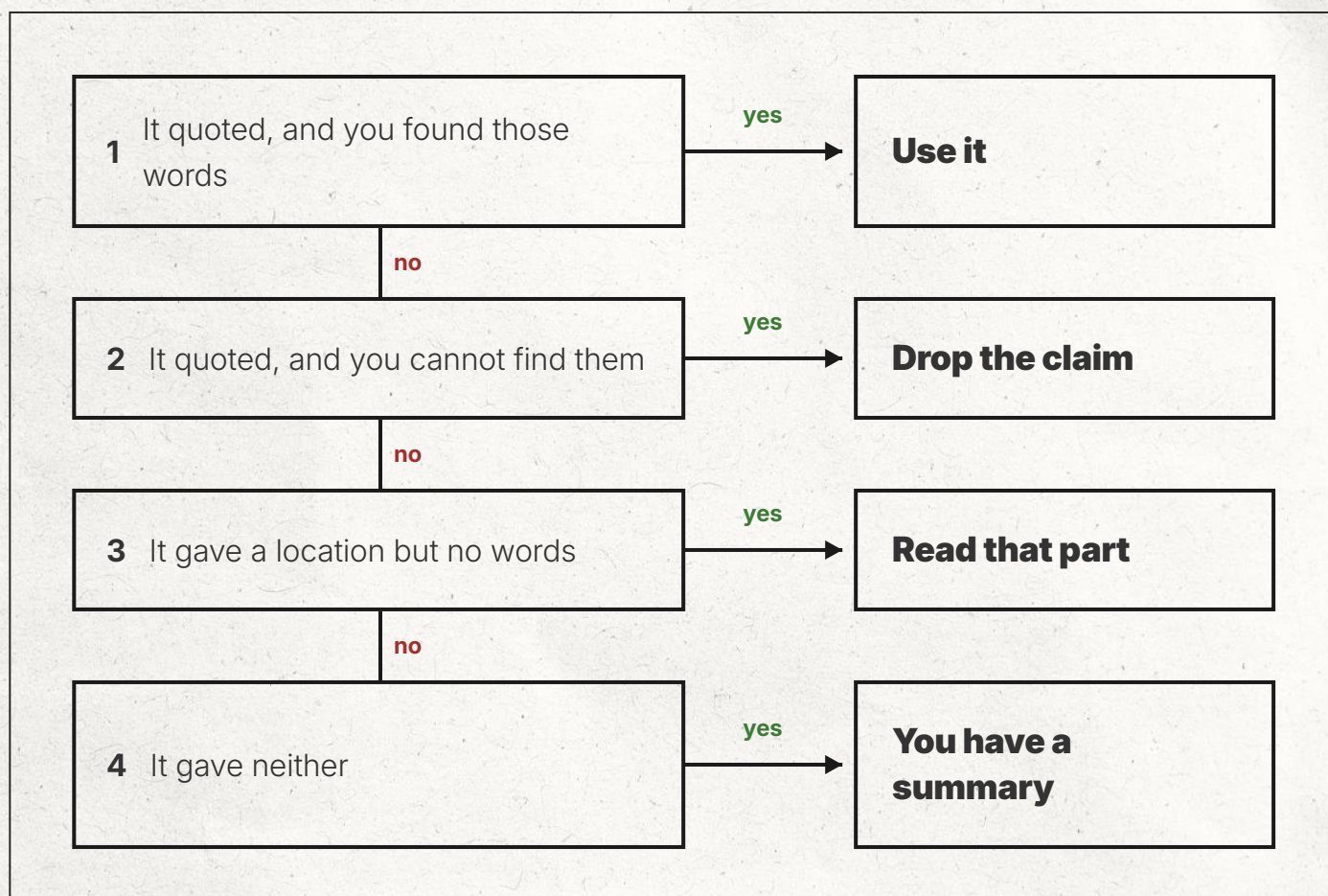


Fig 18.4 Read down until a line is true. The whole check costs seconds, because a string of words either appears in the document or it does not.

18.4

Quote, then search. Copy 6 or 7 words out of the quote and search the original document for them.

Split and count. Ask over each half separately, then compare the two counts against the single pass.

Finding and inventing are separate. Being good at locating real material does not make a model unwilling to produce material that is not there.

The rule is one line long: every claim arrives with the words it came from and where they sit, in the same message. Put it in the instruction itself. If a claim cannot be supported by a passage from the document, say so instead of stating it.

This works for a reason worth knowing. Evaluation work this year found that finding real material and refusing to invent material are separate abilities in these models, not one ability. Two models scored within a point of each other on locating facts that genuinely were in a source, and differed by more than 40 points in how often they stated facts that were not. Competence and honesty are not the same axis, so you cannot infer one from the other, and a fluent answer tells you nothing.

What you can do is check. Copy 6 or 7 words out of a quote it gave you and search the original document for that string. Extracting text verbatim is one of the things these tools do most reliably, error rates in published tests running in the low single digits, which is exactly what makes a failed search meaningful. If the words are not there, the claim resting on them goes. No argument, no second opinion.

That catches invention. The opposite failure, the one that costs more, is the thing it never mentioned at all, and it needs a different check. Ask for every instance of something as a numbered list with locations. Then ask the same question again over the first half of the document only, then over the second half only. If the two halves add up to more than the single pass found, the single pass missed something, and section 18.1 tells you where to expect it to have gone missing.

Both checks together cost you two extra messages and a search box. You are not going to do this for every document you read, and you are not meant to. Do it for the ones you are going to act on, sign, forward to a board, or repeat out loud in a room where somebody knows the document better than you do.

Take the document you brought to this chapter, run question 2 on it, and check one of the quotes. That single check will tell you more about how far to trust this tool than any amount of reading about it.

READING ENDS HERE

Everything before was reading. Everything after is doing.

3 exercises, About 20 minutes and one real decision that opens Chapter 19.

Do these before Chapter 19

3 exercises. About 20 minutes

Take the long document you brought. Before you ask it anything, write down the one thing you most need to know is true or false in that document, and roughly where you expect it to sit.

Run questions 2 and 4 from section 18.3 on your document, as two separate messages. Note what came back in each.

CLAIMS, QUOTED, WITH LOCATIONS

WHAT IS MISSING THAT SHOULD BE THERE

Pick one claim it gave you. Copy 6 or 7 words out of its quote, search the original document for that string, and write down what happened.

Chapter 18 on one page

THE 6 QUESTIONS

What is this and who wants what · the main claims, quoted, with locations · supported or merely asserted · what a document like this usually contains and this one does not · where it contradicts itself · the one thing to act on, and what would make that wrong

CAPACITY IS NOT RECALL

Published limits say what it will accept. Attention is strongest at the start and end of a document, weakest through the middle, and thins as the document gets longer even when nothing else changes

QUOTE OR DROP IT

Require verbatim words plus a location with every claim. Copy 6 or 7 words, search the original. Not found means not used

SPLIT AND COUNT

Ask for a numbered list over the whole document, then over each half separately. If the halves sum higher, the single pass missed something

NEEDED BEFORE CHAPTER 19

One document you have interrogated rather than summarised, and one quote you checked against the original

BRING TO CHAPTER 19

The document you interrogated and the quote you checked. Chapter 19 keeps that same document open and turns to the figures in it.

BOTTOM LINE

You can question a long document instead of compressing it, and prove in seconds whether an answer came out of the document or out of the model.

NEXT CHAPTER

Why AI Gets Numbers Wrong

Chapter 19 stays with your document and turns to the tables inside it. Where an assistant is genuinely reliable with figures, where it produces arithmetic that looks right and is not, and how to tell those apart before a number reaches a slide with your name on it.

BRING WITH YOU

- ☐ The long document you interrogated in this chapter
- ☐ One table or set of figures from it that has to be right

GO FURTHER

The Ultimate AI Library

Every guide, prompt pack and playbook AI Central has published, in one place. The reference shelf this course keeps pointing you to.

explore the library



WHO MAKES THIS



AI Central Media is a London-based editorial team. Our flagship publication covers practical AI for senior professionals, and 100+ AI companies, SaaS platforms and growth teams have advertised with us since 2023.

100+

AI companies, SaaS platforms, education brands and growth teams have advertised with us since 2023

613K

Accounts reached every month, across the 7 channels we operate

151

Countries our readers are in, plus all 50 US states

PROUD PARTNER OF

GAMMA

II ElevenLabs

guidde.

Outskill

Luma AI

HubSpot

attio

S Synthflow

Notion

Replit

We turn attention into pipeline.

visit thecentral.ai

