USE IT IN YOUR WORK

AI at Work

The tasks it genuinely helps with, the ones it quietly ruins, and the line you must never cross.

First edition, September 2026

Addaly

Series editor: Dr Mustafa Mun

Draft, open for academic review

USE IT IN YOUR WORK

AI at Work

The tasks it genuinely helps with, the ones it quietly ruins, and the line you must never cross.

Series editor: Dr Mustafa Mun

Addaly

First edition, September 2026 · Draft, open for academic review

AI at Work

First edition, September 2026, build 62a26075.

Written by AI models to a written standard, with Dr Mustafa Mun as series editor. Exam keys checked blind by a second model: 3,665 of 3,666 agreed. Academic review invited.

The manuscript was drafted with the assistance of a large language model and was edited and published under his name. That is stated plainly here rather than left to be discovered, because a reader is entitled to know how a book they are trusting was made.

How to cite: *Mun, M. (2026). AI at Work. Addaly. addaly.com/learn/ai-at-work.*

This book is the printed form of the course of the same name at addaly.com/learn/ai-at-work. The website is the living version and is corrected first. Where the two differ, the website is right.

Free to read, free to download, free to print, and free to teach from. There is no paid edition and there never will be.

Every diagram in it is drawn from data by code, not generated as a picture, so the numbers on the page are the numbers in the argument. The answers to every check, test and examination question are at the back.

8 modules · 73 lessons · 27 figures · 8 case studies · 58,478 words · about 11 hours

Brief contents

1

MODULE 1

Deciding what to hand over

Audit a week of your own work and place each recurring task into hand over, assist or never, defending each placement by where the truth lives, what a wrong answer costs, and what checking costs.

Before any prompt, there is a decision: which of the things you do this week should touch AI at all. This block gives you a sorting rule you can apply to your own job, the evidence on where the help is real and where it reverses, and the mechanism behind the failure that keeps catching careful professionals.

Lesson 1 · 8 min

What it is actually good at

THE ONE THING TO KEEP

If you did not supply the facts, treat every fact it returns as unverified — fluency is not evidence.

Most advice about AI at work is either "it will replace you" or "it makes things up, ignore it." Both are useless when you have a job to do on Tuesday. You need a sharper question: which parts of my work does this help with, and which parts does it quietly damage?

Here is the most useful distinction I know. It is not about topic. It is about where the truth lives.

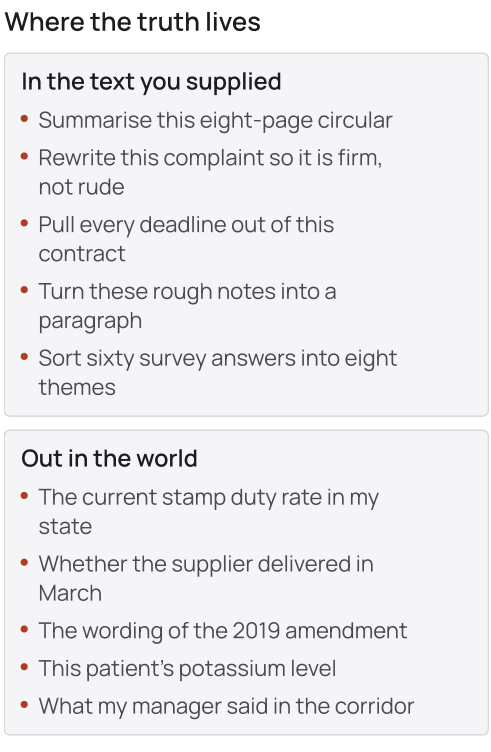
Truth in the text vs truth in the world

When the answer is **inside the text you provided**, the model is strong. Summarise this eight-page circular. Rewrite this complaint letter so it is firm but not rude. Pull every deadline out of this contract. Turn these rough notes into a paragraph. The material is right there. The model is reorganising, compressing, rephrasing. It has everything it needs.

When the answer lives **out in the world**, the model is guessing from patterns. What is the current stamp duty rate in my state? Did this supplier actually deliver in March? What was the exact wording of the 2019 amendment? What is my patient's potassium level? None of that is in the conversation. The model will still answer, fluently and in full sentences, because producing plausible text is what it does. That fluency is the danger. A wrong answer does not arrive looking wrong.

So the first sorting rule: **if you did not give it the facts, do not trust the facts it gives back.**

Figure 1.1



Same model, same fluent sentences, opposite reliability. On the left it is reorganising material it can see; on the right it is guessing, and the guess arrives in the identical register.

Where it earns its keep

Across very different jobs, the same shapes of task keep paying off:

- **The blank page.** A first draft of anything. A parent letter, a tender response, a policy note, a product description. Drafts are cheap to fix and expensive to start.
- **Compression.** Forty pages down to one. A two-hour meeting down to seven decisions.
- **Translation between registers.** Legalese into plain language for a client. Clinical notes into something a family understands. English into Spanish, Hindi, Yoruba, Tagalog — good, not perfect, and better if you can read the output.

- **Structure from mess.** Sixty open-ended survey answers sorted into eight themes. A pile of receipts turned into a table.
- **The thing you can nearly do.** A spreadsheet formula, a regex, a SQL query, a bit of Excel VBA. You cannot write it, but you can test whether it worked.

Where it will hurt you

- **Anything where being 95% right is a failure.** Dosages. Tax filings. Court deadlines. Structural loads. Not "never use it" — never use it *unchecked*.
- **Recent and local specifics.** Last month's circular from your ministry. Your state's fee schedule. Your company's leave policy. Unless you paste it in.
- **Counting and exact arithmetic across long documents.** "How many invoices are over 50,000 naira" is a spreadsheet question, not a chat question.
- **Judgment that is yours to make.** Should we fire this person. Should this student repeat the year. Is this client lying. It will give you an answer. That answer carries no accountability, and yours does.

Do this today

Write down the five things you did last week that took over thirty minutes and were mostly typing or mostly reading. That list is your real curriculum. Most people find two or three of them are pure text-shuffling, and those are the ones to hand over first.

BEFORE YOU MOVE ON

A district health officer needs two things done: condense a 60-page WHO guidance PDF she has open into a one-page briefing, and confirm the current national reimbursement rate for a particular vaccine. Which task is the riskier one to hand to a chatbot?

A. The reimbursement rate, because the number is not in anything she gave the model, so a confident answer may be invented
B. The 60-page summary, because long documents are where models lose accuracy and start hallucinating
C. Both equally, since the model has no way to verify anything it says in either case
D. The reimbursement rate, because health policy is a regulated domain and models are restricted from answering it

The answer and why: p. 409

Lesson 2 · 9 min

What it is actually doing when it answers you

THE ONE THING TO KEEP

The model only predicts the next fragment of text, so it has no memory, no clock, no calculator and no register of what it does not know — and every characteristic failure follows from one of those absences.

The only operation there is

A large language model does one thing. Given a stretch of text, it works out a probability for every possible next fragment, picks one, adds it to the end, and runs the whole calculation again. That loop, a few hundred times over, is the letter it drafted for you.

Nothing else is happening. There is no separate part that checks facts, no lookup step, no moment where it decides whether it knows. Understanding this is not academic curiosity. It tells you precisely which mistakes to expect, and it is the difference between a person who is surprised every week and a person who is never surprised.

It does not see letters

The model does not read characters. Text is first cut into **tokens** — common fragments learned from the training data. In English, a token averages about four characters, so 1,000 tokens is roughly 750 words. Common words are single tokens; unusual ones split. "Unhelpfulness" is likely three or four pieces.

Two consequences you will meet.

The first is the counting failure. Ask how many times the letter "r" appears in "strawberry" and a capable model may say two. It is not being careless. The word arrived as two or three opaque fragments, and the letters inside them were never separately visible. The same model can analyse a lease correctly and miscount a word, because those are different kinds of task on different objects.

The second matters more for cost. Tokenisers were fitted mostly on English text. The same sentence in Hindi, Tamil, Arabic or Amharic can consume two to four times as many tokens as its English translation, because the script fragments into smaller pieces. Since you are billed per token and the context limit is counted in tokens, working in your own language can cost several times more and fill the window several times faster for identical meaning. This is a real inequity, it is rarely mentioned, and it is worth knowing before you conclude that the tool is worse in your language than in English.

No memory, no clock, no calculator

Three absences follow directly from the mechanism.

No memory. Each conversation begins with nothing. When a product appears to remember you, a separate system has saved notes and is quietly pasting them back in at the top of every conversation. That is a feature bolted on, not a property of the model, and it can be switched off — which is worth knowing when the notes contain a client's name.

No clock. It has no access to the time unless the date is written into the hidden instructions the product sends with your message. Most consumer products do send it. Many do not. If you ask "is this contract still within its notice period" without stating today's date, you may be getting arithmetic based on a date it guessed.

No calculator. Arithmetic is produced the same way as prose — by completion. Small sums it has effectively memorised. Multiply two seven-digit numbers and accuracy collapses, in the same confident tone. Modern products often detect a sum and hand it to a real calculator or a snippet of code, which fixes it. Whether yours does is something you should test with a multiplication you can verify, not assume.

Where its knowledge sits

What the model knows is stored in its weights — billions of numbers fixed at the end of training. It is not a database. There is no row to inspect, no source to click, no field that says "confidence". You cannot ask it what it does not know, because there is no register of that anywhere in the system.

This is why the honest instruction throughout this course is that facts you did not supply are unverified. Not usually wrong. Unverified, which is a different and more useful category.

The same question, a different answer

Ask the same question twice and you often get two different replies. That is sampling: it is drawing from a distribution rather than reading out a stored answer. Some tools let you set a "temperature" of zero to take the most likely

fragment every time, which reduces variation but does not eliminate it — the hardware itself is not perfectly reproducible when work is batched across users.

You can turn this into a tool rather than an annoyance, and a later lesson does exactly that: where two runs disagree is where the model is least certain, and that is where you look.

Watch it happen, for nothing

The clearest way to internalise all of this is to run a small model yourself. **Ollama** (free, Windows, macOS, Linux) will install and run a two- or three-billion-parameter model on an ordinary laptop with 8 GB of memory. `llama.cpp` does the same with less software around it. **Hugging Face** hosts free tokeniser demos in the browser, where you can paste a sentence and watch it cut into pieces — try one line of English and the same line in your own language, and read the two counts.

A small model fails in the same ways as a large one, only more often and more visibly. An hour with one teaches more about the shape of the failures than a month of reading about them.

BEFORE YOU MOVE ON

A model correctly analyses a twelve-page lease, then says the word "strawberry" contains two letter r. What does this pair of results tell you?

A. Text reaches the model as tokens rather than characters, so letter-level questions ask about something it never received, while sentence-level meaning is exactly what it does receive.

B. The model was being lazy on the easy question and would have got it right if asked to think step by step.

C. Counting is arithmetic, and the model has no calculator, so it fails at all numeric questions equally.

D. The lease analysis is probably also unreliable, since a system that cannot count letters cannot be trusted with legal text.

The answer and why: p. 410

Lesson 3 · 9 min

The context window, and why a long chat gets worse

THE ONE THING TO KEEP

Every turn resends the entire conversation, so instructions given early can be truncated away, attention is weakest in the middle, and cost grows with the square of the chat's length — which is why one task should mean one conversation.

Everything it knows, every time

When you send your fortieth message in a conversation, the model does not remember the previous thirty-nine. It is handed all of them again, as one long block of text, and reads the lot from scratch before writing a word.

CASE STUDY · MODULE 1

Twelve seats, or one stopwatch

An illustrative case, built from documented patterns.

Sunrise Academy, a coaching centre in Kota with eleven teachers, about four hundred students and a director who has just been sold a subscription.

Vikram Sethi came back from a franchise meeting holding a quotation for twelve seats on an AI assistant at roughly two thousand four hundred rupees a seat a month. Annualised, that is a little under three and a half lakh, which at Sunrise is one junior teacher.

He had two tasks in mind, and he was clear about which one he wanted.

The first was the weekly parent letter. Forty of them, one per batch, mostly the same: what was covered, what the test dates are, who has fallen behind on attendance. A senior teacher writes them on Saturday evening and complains about it every Saturday evening.

The second was the one that had sold him the idea. After every fortnightly test the centre sends each family a short note explaining why their child's rank moved. Four hundred notes, twice a month, written by whichever teacher can be cornered. Everyone hates them and parents read every word. If the assistant could write those, Sunrise would get a Sunday back and, Vikram thought, would look more professional than the two coaching centres on the same road.

His operations head, Anjali Rao, asked him to hold the purchase order for two weeks and answer three questions about each task instead.

Where does the truth live? For the parent letter, in the syllabus tracker and the attendance register — documents Sunrise holds and can paste in. For the rank note, the arithmetic of the rank is in the test data, but the *reason* is not. The reason is that the boy missed nine days with dengue, that the girl has been put in a batch that moves faster than she does, that a whole cohort was thrown by a badly set question in section B. None of that is in any file. It is in the teachers' heads and in the corridor.

What does a wrong answer cost? A clumsy sentence in a weekly letter costs a shrug. A wrong explanation of why a child's rank fell costs a meeting with a father who has just been told his daughter is not working hard enough, when in fact she was ill and the centre knew it.

What does checking cost? Here the two tasks separate completely. A parent letter is checked by reading it, which takes ninety seconds and is something the teacher would have done anyway. A generated rank explanation cannot be checked by reading it, because a plausible reason and the real reason look identical on the page. Checking one means going to the attendance register, the batch record and the teacher — which is longer than writing the note from those things in the first place.

That left Vikram with a decision he did not like. Buying twelve seats and starting with the rank notes would take the most hated job off his staff immediately, and it was the job that had made the tool look worth paying for. Refusing to start there meant telling eleven people that the thing they were promised was not coming, and one of his best teachers had already said she would not do another term of Sunday nights.

The other side of it was not hypothetical either. Sunrise's whole business is that parents believe the centre knows their child. A single note explaining a rank drop by a reason the centre had invented would be worth more damage than a year of Sunday evenings.

YOUR ANSWERS

1. Anjali's objection to the rank notes was not that the model would write badly. What was it?

2. Sunrise reversed the direction of the rank-note task rather than abandoning it. Why does that reversal make it safe?

3. Sunrise added a rule that a note may not contain a reason absent from the bullets, and it caught an error that no amount of care had caught. What does that tell you about relying on staff vigilance?

STOP HERE

Write your answers before you turn the page. How it played out starts on p. 52.

This page is blank so that how it played out stays out of view until you turn the page.

CASE STUDY · MODULE 1

How it played out

Vikram bought two seats rather than twelve and gave himself a term to decide about the rest. The weekly parent letter went into the hand-over pile: a brief with the syllabus tracker and the attendance extract pasted in, a ninety-second read before sending. Timed over six weeks it fell from about thirty-four minutes to eleven, including the check. The rank note went into the assist pile with the direction fixed the other way round — the teacher writes three bullets of their own judgement, and the model turns them into a paragraph at the right register in the family's preferred language. Sunrise added one rule to the brief: the note may not contain a reason that is not in the bullets. In the second batch a teacher approved a note that explained a drop by "reduced practice at home", which the model had supplied and nobody had said; it was caught by the rule and not by anybody's vigilance, which is why the rule exists. At the end of the term Vikram bought six more seats, not ten, because the audit had found only eight people whose week contained a task that actually moved.

The questions, discussed

1. Anjali's objection to the rank notes was not that the model would write badly. What was it?

That the truth for a rank note lives outside any document Sunrise could supply. The model can compute nothing about why a rank moved, so it produces the reason such notes usually give — more practice, less distraction, exam nerves — in exactly the register a good teacher would use. The failure is invisible on the page, and checking it means consulting the attendance register, the batch record and the teacher, which is the work the note was supposed to save.

2. Sunrise reversed the direction of the rank-note task rather than abandoning it. Why does that reversal make it safe?

Because it moves the judgement to the person who has it and leaves the model only the phrasing. The teacher supplies the cause in three bullets; the model expands them. Every fact in the finished note came from somebody who knew

it, so there is nothing for the model to invent, and the checking cost drops from consulting three records to reading a paragraph against the bullets you just wrote.

3. Sunrise added a rule that a note may not contain a reason absent from the bullets, and it caught an error that no amount of care had caught. What does that tell you about relying on staff vigilance?

That vigilance degrades exactly where the tool is usually right. A teacher reading a fluent, plausible note about their own pupil has no signal that one clause was supplied rather than reported. A mechanical rule — nothing in the output that was not in the input — does not depend on anyone feeling alert on a particular evening, which is the only kind of check that survives a busy fortnight.

MODULE 1 TEST

Check what stuck

6 questions, one right answer each. This one is for you, not for a record. The answers, and the reason behind each, are in the key on p. 408. If two go wrong, re-read the module before the exam.

1 You attach an eight-page circular and ask for a summary. In a separate chat with nothing attached, you ask for your state's current stamp duty rate. Both replies are equally fluent. Why is only one of them cheap to rely on?

A. Tax and legal questions sit outside the material these models were trained on, whereas ordinary administrative prose is very well represented in it

B. You supplied the facts in one case and not the other, so only one reply can be checked against something in front of you

C. The summary is longer, and extra length gives the model more opportunity to correct itself while it is writing

D. Rates change frequently, so the model declines to answer and produces a hedged approximation in place of a figure

2 A colleague finds that the same paid tool costs her roughly three times as much per document in Tamil as in English, for documents that say the same thing. What is going on?

A. The provider applies machine translation before and after every request, in both directions, and charges for each of those additional passes

B. A smaller model is used for languages other than English, so more attempts are needed to reach the same answer

C. Her script fragments into far more tokens for the same meaning, and both billing and the context limit are counted in tokens

D. Documents in her language genuinely contain more characters, because the words used in them are longer on average

3 Twenty messages into a long conversation, the model stops obeying an instruction you gave carefully at the start. What has most likely happened, and what is the cheapest fix?

A. The instruction was too vague to act on, and rewriting it as a longer and more formal rule will make it hold

B. The conversation has passed a usage threshold and the product has switched you to a smaller model for the rest of the session

C. The model has judged the instruction no longer relevant, and it needs to be told more firmly that it still applies

D. The early part has been truncated or summarised away, so restate the requirement in your next message

4 A solicitor asks for a summary of a forty-page expert report, gets a good one in ninety seconds, and then reads all forty pages anyway because she will rely on it at a hearing. What is the right change to make?

A. Ask instead for which paragraphs discuss causation, quantum and the claimant's history, so wrong entries show up on opening them

B. Ask for a much longer summary, so that less is left out and there is less of the report left to reread

C. Ask the model to confirm that nothing material has been omitted, and rely on the summary once it has confirmed

D. Use a reasoning model instead, since the extra deliberation makes it substantially more reliable at noticing what a summary has left out

5 Sixteen experienced developers expected AI help to make them about a quarter faster, reported afterwards that it had made them about a fifth faster, and were measured as about a fifth slower. What should you take from this about your own work?

A. These tools slow experienced people down, so they should be reserved for junior colleagues and new starters
B. Self-reports are reliable about speed but not about quality, so quality needs to be measured separately
C. Do not trust your sense of your own speed; time the task instead, with the checking inside the timing
D. Gains only appear on unfamiliar work, so anyone doing their core craft should not use these tools at all

6 Your team's answer to automation bias is that a second colleague reads every generated draft before it goes out. Why does that do much less than it appears to?

A. Second reviewers read faster than first reviewers, and speed is what causes most missed errors in practice
B. The second reviewer usually has less domain expertise, so their reading adds little beyond a spelling check
C. Reviews of generated text are less thorough because people assume the first reviewer has already checked the figures
D. Both readers are anchored by the same text in the same order, so the same omissions are invisible to both

Figure 2.1

Five parts of a brief

The situation, in your own words
What has happened, who is involved, what has already been said. Three sentences of context beat three paragraphs of instruction.

The reader
Not a professional audience. A district officer who gets forty of these a week. A parent who is already angry. A client who does not know what an indemnity is.

The constraint
Under 120 words. One side of A4. Must include the refund date. Do not apologise, we did nothing wrong.

What good looks like
They reply with a date and do not ask what we mean by in principle. One line, and it converts a writing task into a purpose.

What bad looks like
No hoping this finds you well, no reaching out, nothing that sounds like a form letter. Naming a banned phrase works because it is concrete.

Two or three of these lift the draft more than any list of style adjectives, because each one supplies a fact about your situation that the model could not have guessed.

A worked pair

Weak: *Write a polite but firm reminder about an overdue invoice.*

Better:

> *A client of six years is 47 days late on a ₹2,80,000 invoice. Their accounts contact has not replied to two emails. I know the director personally and want to keep the relationship. This is the third reminder and should be the*

> *last before I involve our accountant. Under 150 words, no threats, one clear ask with a date. It goes to the director, not to accounts.*

Same task. The second one produces something you might send after two edits, because everything the writer knew is now in the room.

Iterate on one thing

When the draft is not right, most people rewrite the whole prompt and change six things at once. Then the next one is worse in a different way and they conclude the tool is unreliable.

Change **one** thing and look. "Too formal" is not an instruction; "cut the first paragraph and start with the ask" is. "Make it better" is not an instruction; "same content, 90 words" is. Editing a draft in front of you is far more effective than re-describing what you wanted from scratch, because the draft is now context too.

And when three attempts have not got there, stop prompting and start typing. A draft that is 80% right is a starting point, not a negotiation. The moment you are arguing with it, you have crossed the line where doing it yourself is faster — which is the verification-cost trap from the first module, wearing different clothes.

Where the brief goes

Once you have written a good brief for a recurring document, you have made an asset, not a message. Save it. Replace the specifics with blanks. Next month, fill in the blanks.

That is the whole idea behind saved prompts, custom instructions and the "project" features in the major tools, and it is covered properly in the last module. For now, keep a plain text file. Six good briefs in a text file will do more for your week than any subscription.

EXAMINATION

AI at Work — final examination

This paper covers the whole course:

- deciding which of your tasks to hand over
- briefing and editing text you would sign
- decomposing work that is too large for one prompt
- working from your own documents and data
- the tools already on your desk
- the confidentiality and legal lines you must not cross
- how the method applies inside your own trade
- how to make any of it hold

56 questions, the whole pool. The site draws 25 for a sitting, pass mark 70%. Work any 25 under your own conditions. Answers from page 457.

1 Module 1 · Deciding what to hand over

A brief you are drafting needs five supporting examples. Only four exist. What does asking for exactly five do, and what is the better instruction?

A. It causes the model to widen its search and return four with a note explaining that a fifth could not be found in the material

B. It applies pressure to produce five, so the fifth is invented; ask for up to five and for it to say if there are fewer

C. It has no effect on invention, because the count is a formatting instruction rather than a content one, so ask for any number you like

D. It makes the model reuse one of the four in different words, which is easy to spot by reading the list carefully

2 Module 1 · Deciding what to hand over

Where does fabrication concentrate, and why there rather than in the argument?

A. In long passages, because attention thins as generation continues and the later sentences are less well grounded in the material than the earlier ones

B. In technical vocabulary, because specialist terms appear less often in training data than ordinary words do

C. In the opening paragraph, because the model commits to a position before it has read the whole of the material you supplied

D. In identifiers, precise figures attributed to a source, and named local specifics, because those are highly patterned and must be produced exactly

3 Module 1 · Deciding what to hand over

Four right and one wrong is described as the dangerous shape of an answer. Why is it more dangerous than one that is wholly wrong?

A. Because a mostly correct answer takes longer to check in full, and people abandon a check that has already taken several minutes

B. Because errors in a partly correct answer are usually in the middle, and the middle is where attention is weakest in any document

C. Because the four correct entries buy your trust for the fifth, and nothing in the fifth looks different from the others

D. Because a wholly wrong answer would be caught by the tool's own safety checks before it ever reached you on screen

4 Module 1 · Deciding what to hand over

Which task sits in the free-to-check band, and what makes the band a property of the task rather than of the tool?

A. A regular expression tested against strings you already know the answers for; the answer verifies itself against something you hold

B. A summary of current data protection requirements in your country, because the requirements are published and can be looked up by anybody

C. A translation into a language you do not read, because a second tool can be used to translate the result back again

D. An account of what was agreed in a meeting you did not attend, because a colleague who was there can confirm it afterwards

5 Module 1 · Deciding what to hand over

Why is AI least safe at the edge of your competence, which is exactly where it feels most valuable?

A. Because models are trained mainly on introductory material, so their answers degrade sharply on anything specialist

B. Because you cannot check what you could not have written, and omission is invisible by definition

C. Because unfamiliar subjects require longer prompts, and long prompts dilute each instruction they contain

D. Because a topic you know less about is one you will spend less time reading, so the check is shorter than it should be

6 Module 1 · Deciding what to hand over

You paste 300 rows of sales into a chat window and ask for the total of one column. Which product shape is the right one, and why?

A. The chat window is correct for this, provided you ask it to work through the addition in steps and to show its arithmetic before the total

B. The transcriber, because it is built to process structured input at volume rather than to converse about it

C. The in-suite assistant, because it can see the underlying file rather than the version you pasted into a conversation

D. The code-running mode, or a spreadsheet formula, because those do arithmetic instead of predicting text that resembles arithmetic

Module 1 · Deciding what to hand over

The module starts on p. 11.

MODULE 1 TEST

Module 1 test, Q1, p. 54
You attach an eight-page circular and ask for a summary. In a separate chat with nothing attached, you ask for your state's current stamp duty…
Answer B: You supplied the facts in one case and not the other, so only one reply can be checked against something in front of you
Why: The cost of checking is a property of where the facts came from, not of the topic. With the circular attached, every claim is verifiable against a page you are holding: cheap. With nothing attached, every claim requires research you have not done: expensive. Same model, same fluency, an hour of difference in what it costs you to be responsible for the answer.

Module 1 test, Q2, p. 54
A colleague finds that the same paid tool costs her roughly three times as much per document in Tamil as in English, for documents that…
Answer C: Her script fragments into far more tokens for the same meaning, and both billing and the context limit are counted in tokens
Why: Tokenisers were fitted mostly on English text, so the same sentence in Tamil, Hindi, Arabic or Amharic can consume two to four times as many tokens. You are billed per token and the window is measured in tokens, so working in your own language costs more and fills the window faster for identical meaning. It is a real inequity and it is rarely mentioned.

Module 1 test, Q3, p. 55
Twenty messages into a long conversation, the model stops obeying an instruction you gave carefully at the start. What has most likely happened, and what…
Answer D: The early part has been truncated or summarised away, so restate the requirement in your next message
Why: Nothing outside the context window exists. When a conversation overflows, some products silently drop the oldest messages, so the instruction is no longer in front of the model and nothing remains that knows it was ever given. Repeating it late puts it at the end of the window, where attention is strongest. Scolding carries no information and adds an argument to the context.

Module 1 test, Q4, p. 55
A solicitor asks for a summary of a forty-page expert report, gets a good one in ninety seconds, and then reads all forty pages anyway…
Answer A: Ask instead for which paragraphs discuss causation, quantum and the claimant's history, so wrong entries show up on opening them
Why: This is a task where verification cost exceeds production cost, and no improvement in the model fixes that: if you must check every line against the source, you have to read the source. Splitting the task rescues it. A navigation aid is checked by opening the paragraph, which takes seconds, so the value is real and the reading stays hers.

Module 1 test, Q5, p. 56

Sixteen experienced developers expected AI help to make them about a quarter faster, reported afterwards that it had made them about a fifth faster, and…

Answer C: Do not trust your sense of your own speed; time the task instead, with the checking inside the timing

Why: The finding is about calibration, not about a verdict on the tools. Waiting for a draft feels like progress, and the correcting and re-prompting fragments into pieces too small to register as work, so the felt saving and the measured one come apart by roughly forty points in the flattering direction. A phone stopwatch settles in ten minutes what surveys cannot settle in a year.

Module 1 test, Q6, p. 56

Your team's answer to automation bias is that a second colleague reads every generated draft before it goes out. Why does that do much less…

Answer D: Both readers are anchored by the same text in the same order, so the same omissions are invisible to both

Why: Independence is what makes a second check valuable, and reading somebody else's finished prose destroys it. Two people reviewing one draft are two people led down the same path with the same gaps unmarked. A real second check gives the reviewer the source — the transcript, the file, the original figures — and asks for their answer, not their assessment of this answer.

THE LESSON CHECKS

Lesson 1, p. 15

A district health officer needs two things done: condense a 60-page WHO guidance PDF she has open into a one-page briefing, and confirm the current…

Answer A: The reimbursement rate, because the number is not in anything she gave the model, so a confident answer may be invented

Why: If you did not supply the facts, treat every fact it returns as unverified — fluency is not evidence.

B The 60-page summary, because long documents…: Believes length itself causes hallucination, when the real risk factor is whether the source material was supplied at all.

C Both equally, since the model has…: Treats all model output as uniformly unreliable, which throws away the genuine difference between reorganising given text and recalling world facts.

D The reimbursement rate, because health policy…: Attributes the risk to subject-matter sensitivity rather than to the absence of source material.

Lesson 2, p. 19

A model correctly analyses a twelve-page lease, then says the word "strawberry" contains two letter r. What does this pair of results tell you?

Answer A: Text reaches the model as tokens rather than characters, so letter-level questions ask about something it never received, while sentence-level meaning is exactly what it does receive.

Why: The model only predicts the next fragment of text, so it has no memory, no clock, no calculator and no register of what it does not know — and every characteristic failure follows from one of those absences.

B The model was being lazy on…: Assumes effort is the missing ingredient, when the information needed to answer was never present in the input at all.

C Counting is arithmetic, and the model…: Confuses two separate absences: arithmetic weakness is real, but this particular failure is about tokenisation hiding characters, not about sums.

D The lease analysis is probably also…: Treats capability as a single dial, when the two tasks operate on different objects: the lease arrives as meaningful text, the letters inside a token are not separately visible at all.

Lesson 3, p. 23

At the start of a long chat you instructed the tool never to name the client. Thirty messages later it names the client. What is…

Answer D: The early messages were dropped or summarised to keep the transcript inside the context window, so the instruction is no longer in front of the model at all.

Why: Every turn resends the entire conversation, so instructions given early can be truncated away, attention is weakest in the middle, and cost grows with the square of the chat's length — which is why one task should mean one conversation.

A Model quality degrades over a session…: Imagines fatigue in a stateless system; the weights are identical on turn one and turn thirty, only the text supplied has changed.

B The model decided the instruction no…: Attributes an intention to a system that has no continuous state between turns; there was no decision, only text that was or was not present.

C Safety filters override user instructions after…: Invents a mechanism; nothing in these products relaxes a user's formatting or confidentiality instruction as a function of turn count.

USE IT IN YOUR WORK

AI at Work

The tasks it genuinely helps with, the ones it quietly ruins, and the line you must never cross.

WHAT IT TEACHES	Eight modules and seventy-three lessons on using AI in a real job, without pretending it is magic and without pretending it is useless. Written for accountants, teachers, nurses, lawyers, shopkeepers, marketers and civil servants anywhere in the world.
WHAT IS INSIDE	Eight modules and 73 lessons, 27 figures drawn from data, eight case studies, eight module tests, a 56-question examination, and every answer with the reason behind it.
LICENCE	CC BY-NC-SA 4.0. Copy, print, share and adapt it for any non-commercial purpose, with credit, under the same licence. There is no paid edition.
THE COURSE	Free to read, with the lessons, the films and the examination, at **addaly.com/learn/ai-at-work**

Scan to open the course

Addaly

First edition, September 2026 · build 62a26075