

USE IT IN YOUR WORK

AI for Teachers

For teachers who now have to teach and use AI, often with no training and no projector.

First edition, September 2026

Addaly

Series editor: Dr Mustafa Mun

Draft, open for academic review

USE IT IN YOUR WORK

AI for Teachers

For teachers who now have to teach and use AI, often with no training and no projector.

Series editor: Dr Mustafa Mun

Addaly

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

AI for Teachers

© 2026 Addaly.

Licensed under Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0). You may copy, print, share and adapt it for any non-commercial purpose, with credit, under the same licence.
creativecommons.org/licenses/by-nc-sa/4.0

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 for Teachers. Addaly. addaly.com/learn/ai-for-teachers.*

This book is the printed form of the course of the same name at addaly.com/learn/ai-for-teachers. 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.

7 modules · 57 lessons · 26 figures · 7 case studies · 44,923 words · about 8 hours

Brief contents

	Contents	6
1	Understanding the machine you are about to use	9
2	Prompting as a teacher	49
3	Planning and making materials	89
4	Assessment, feedback and marking	133
5	Integrity: AI-written work and what to do about it	173
6	Teaching AI itself	211
7	The multilingual, mixed-ability classroom	253
	Examination	293
	Answers	313

1

MODULE 1

Understanding the machine you are about to use

Explain to a class, a colleague and a parent what a language model does and does not do, predict from the mechanism which kinds of question it will answer reliably and which it will get confidently wrong, and choose a free tool that runs on the phone you actually have with its training-on-your-data setting turned off.

1	Explaining AI to a class that argues back	10
2	The free tools you actually have	13
3	Where it is reliable, and where it is not	17
4	Why it cannot look it up	21
5	The conversation has a memory limit	26
6	It has read the American curriculum	30
7	A chatbot is not a calculator	33
8	Same question, different answer	37
	<i>Case study: The tool everyone already had</i>	41
	<i>Module 1 test</i>	46

Lesson 1 · 7 min

Explaining AI to a class that argues back

THE ONE THING TO KEEP

It predicts likely next words from human writing, so sounding right and being right come apart.

The one sentence to start with

"It is a machine that has read an enormous amount of writing by people, and it guesses what words come next."

That is the whole thing. Everything else is elaboration, and this sentence survives the questions children actually ask. "It's a robot brain" does not survive. "It's like a search engine" does not survive — the first follow-up is "then why is it wrong sometimes?" and you have no answer.

Make them do it before you explain it

Write on the board:

The old woman walked slowly into the _____

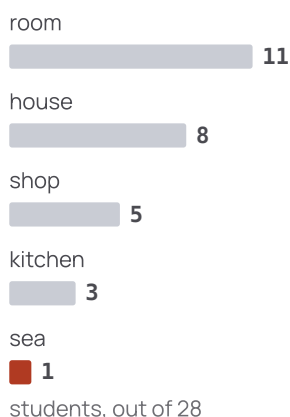
Every student writes one word. Silently, no discussion. Then take a tally by show of hands. Room. House. Shop. Kitchen. Somebody will write "sea" and the class will laugh.

Put the tally on the board. You have just built the thing you are teaching: twenty-eight people, four common answers, one rare one. The machine has that same shape of guess, except it built it from billions of sentences instead of one classroom.

Now the second stem:

The capital of Japan is _____

Figure 1.1

Twenty-eight guesses at one blank

The tally from one Class 7 room for "The old woman walked slowly into the _____". Four common answers, one rare one. The same class, given "The capital of Japan is _____", produced twenty-seven Tokyos and one Kyoto. The scatter is the whole explanation of why the machine is dependable about Tokyo and unreliable about your district.

Nearly unanimous. Tokyo.

Ask them what changed. They will get there: for one question almost everyone's guess is the same, because the writing they have seen nearly always says the same thing. For the other, many guesses are fine. That is the entire reason AI is dependable on some questions and unreliable on others, and a ten-year-old can now explain it to their parents.

The questions they will ask, and honest answers

"Does it know it's me?" No. Usually it starts each conversation with nothing about you. Some products keep notes about you; that is a feature someone built on top, not the machine remembering.

"Is it alive? Does it have feelings?" It writes sentences about feelings because people wrote sentences about feelings and it learned from them. Whether anything is going on inside is genuinely argued about by the people

who build these systems. You are allowed to say that. Try: "Nobody has shown it feels anything, and most researchers think it doesn't. But it is a real argument, and you are old enough to know it's an argument rather than a settled fact."

"Why does it lie?" Lying means knowing the truth and saying something else. It does not do that. It produces the sentence that fits, and a sentence can fit beautifully and be false. It has no way to check. Give them the phrase: it is not lying, it is **making things up without knowing it**.

"Where did it get all the words?" From writing by people — websites, books, articles, forums. That is also why it repeats what people say, including the unfair things people say.

"Can it do maths?" Sometimes, badly, unless a calculator is running underneath. Guessing the next word is a poor way to multiply.

Two analogies to avoid

"It's a robot brain." This invites every film they have seen into the room and you spend the period on robots instead of prediction.

"It's just autocomplete." Closer, and useful, but if you stop there a student will show you something it wrote that is genuinely good and conclude you were wrong. Say instead: "It works like autocomplete, and it turns out autocomplete over enough writing is far more powerful than anyone expected — including the people who built it."

The line worth ending on

"It is very good at sounding right. Sounding right and being right are different things, and telling them apart is now part of your job."

Eleven-year-olds understand that sentence immediately, and it carries the rest of the course.

BEFORE YOU MOVE ON

A student tells the AI it made a mistake. It apologises and changes its answer. She says this proves it understood its error. What should you tell her?

- A. An apology is simply the kind of sentence that follows a correction in the writing it learned from — it will produce one whether or not it was wrong. Tell it a correct answer is wrong and watch it fold.
- B. She is right — noticing and correcting an error requires holding some model of what is true.
- C. The company that built it wrote a rule that it must apologise politely when challenged.
- D. When challenged, it went and looked up the correct answer, which is why it changed.

The answer and why: p. 315

Lesson 2 · 9 min

The free tools you actually have

THE ONE THING TO KEEP

Every major model has a free tier that runs in a phone browser or inside WhatsApp; the choice that matters is not which is cleverest but which one you have turned off training on your data.

You do not need to buy anything

Every major model has a free tier, and every free tier runs in a phone browser. If you have a phone that can open a web page, you have access to the same underlying models a paid user has, with two differences: a cap on how many messages you can send before it slows you down or switches to a smaller model, and fewer extras like long file uploads.

For a teacher generating a worksheet, a set of misconceptions and a parent letter in an evening, the cap almost never bites. Where it does, the fix is to wait, or to have a second tool open.

The realistic list

Here is what is available at no cost as of this writing. Products change names and limits often, so treat the shape as stable and the details as something to re-check.

- **ChatGPT** (OpenAI). Free tier includes the main model with a message cap, image upload, voice, and web search. Works in a browser without the app. Requires an account, which can be a phone number or email.
- **Gemini** (Google). Free with any Google account. Tightly tied to Docs, Sheets and Gmail, which matters if your school runs on Google Workspace. Also inside Android as the assistant.
- **Claude** (Anthropic). Free tier with a daily cap. Strong on long documents and careful writing. Account required; terms are for adults.
- **Copilot** (Microsoft). Free, built into Edge and Windows, uses OpenAI models underneath. If your school runs on Microsoft, this is the one already installed.
- **Meta AI inside WhatsApp**. No new account, no new app, works on the data plan you already pay for. The model is Meta's Llama. It is weaker than the others on long, careful tasks and stronger than all of them on one thing: it is already on the phone of nearly every teacher in India, Nigeria, Kenya and Brazil.
- **DeepSeek** and other Chinese models. Free, capable, and hosted in China, which is a data-location question your school may have a view on.

The open-source route also exists: models like Llama and Qwen can be run on your own computer with free software such as Ollama or LM Studio. This needs a laptop with at least 8 GB of memory and gives you a weaker model in exchange for nothing leaving the machine. It is the right answer for a school that cannot send any data out, and the wrong answer for a teacher with a phone.

The one setting

Most consumer chatbots default to using your conversations to improve their models. That means the essay you pasted, the student name you forgot to remove, the parent complaint you asked for help with — all of it may be read by people at the company and folded into future training.

Before you paste anything about a student, find the setting. In ChatGPT it lives under Settings, then Data controls, labelled something like "Improve the model for everyone". In Gemini it is under Activity. In Claude it is a privacy setting. Turn it off. This does not make the tool private — the company still receives and stores what you type — but it takes your students' work out of the training pipeline.

Then treat what you paste the way you would treat what you say in a staffroom with the door open. First names and no surnames. No addresses. Nothing you would not want a stranger at the company to read.

Browser or app

Use the browser. The apps ask for permissions you do not need to grant, take storage you may not have, and on a shared family phone leave a logged-in account for whoever picks it up next. A browser tab you close is gone. If the phone is shared, use a private window.

The honest limitation

A free tier is a queue. When demand is high, the service quietly gives you a smaller, faster model, and answers get shallower without any notice. If a tool that was sharp at seven in the morning is producing thin work at nine in the evening, that is usually why. Nothing you did changed. Ask again tomorrow, or switch to a second tool for the rest of the session.

The other limit is the message cap itself. It resets on a schedule — typically every few hours — and it is per account, so a department sharing one login will hit it by lunchtime. Each teacher needs their own.

What to do tonight

Open one of these in your phone browser. Turn off training. Ask it for five wrong answers your students give about something you taught this week, and see how many you recognise. That is the whole setup, and it took ten minutes.

BEFORE YOU MOVE ON

A teacher on a shared Android phone with patchy data wants to generate practice questions each evening. Which single setup decision matters most before she starts?

- A. Picking whichever tool scored highest on the latest benchmark leaderboard, because a stronger model will make fewer mistakes on the worksheets she generates.
- B. Turning off the setting that lets the company train on her conversations, and working in the browser.
- C. Installing every app on the phone so she can compare the answers across all of them for each question she asks.
- D. Paying for one subscription first, since free tiers are demonstration versions that are not intended for real professional work.

The answer and why: p. 316

CASE STUDY · MODULE 1

The tool everyone already had

An illustrative case, built from documented patterns.

A government secondary school in Barabanki district, Uttar Pradesh, choosing one free AI tool for forty-one teachers.

The school had forty-one teachers and a computer room with two working machines. Every teacher had a phone. Perhaps half had data that lasted to the end of the month. When the principal asked Sunita Verma, the science head, to recommend one tool for the staff, the question looked like a comparison of models and turned out not to be.

There were two realistic candidates. The first was the assistant inside WhatsApp. It needed no new account, no new app and no email address, it ran on the data plan people already paid for, and the staff group where timetables were shared was already open on every phone. The second was ChatGPT or Gemini in the phone browser: better on long, careful work, better at holding a pasted syllabus, and requiring each teacher to create an account and then walk through a settings menu to find the control that stops conversations being used to train the model.

Mrs Verma tested both for a fortnight, on the work she actually had to do. Most of what she found was reassuring. Both explained photosynthesis and the water cycle well enough to use. Both, asked what Class 8 students get wrong about the water cycle and what they are thinking when they get it wrong, produced eight answers, two of which she had not met in nineteen years. That single request turned out to be the most valuable thing either tool did all fortnight, and it cost nothing but a sentence.

Both also produced a Class 8 worksheet on the district's part in the freedom movement, fluently, with a founding year for the town's oldest school — and the two years did not match. She checked the school's own foundation stone. Neither was right. She noticed, separately, that the answers she got at nine in the evening were thinner than the answers she got at six in the morning, which she assumed was her own tiredness until a colleague reported the same thing.

All of that was useful and none of it was the decision. The decision was the setting.

In the browser tool she had found the control in four minutes and switched it off. In the WhatsApp assistant she looked for twenty minutes, could not find an equivalent, and could not get an answer from anyone at the district office about whether one existed. She knew what teachers in that staff group already did, because she was in it: they forwarded photographs of student scripts to each other and asked for help marking them.

The two costs

Recommending WhatsApp meant something close to full adoption. The eight teachers with the oldest handsets could use it. The five who had never had an email address could use it. It would be in the staffroom by Monday, and the material generated with it — worksheets, explanations, question lists — would be real and useful, because the tool that gets used is the tool people already have open.

It also meant she would be recommending, to forty-one colleagues, a tool she could not tell them was safe to paste a child's work into. She could write a rule. She could not enforce one.

Recommending the browser tool meant a better position on data and a worse one on everything else. Her estimate, from the two afternoons she had spent helping colleagues make accounts, was that perhaps twelve of the forty-one would complete the setup and keep using it. The rest would carry on as before, and the tool would become one more thing that the better-connected half of the staffroom had and the other half did not.

There was a third option she considered and dropped. A laptop in the office could run a small open model with nothing leaving the machine, which answers the data question completely. It also meant one machine, in a room that is locked at four o'clock, running a model noticeably weaker than either candidate, for a staff of forty-one who plan their lessons at night. It is the right answer for a school that cannot send any data out and the wrong answer for a teacher with a phone, and this was a school of teachers with phones.

She had one staff meeting, forty minutes, and a principal who wanted a single answer.

YOUR ANSWERS

1. Both tools gave a founding year, the years differed, and both sounded equally certain. Explain why, from the mechanism.

2. Her rule depended on teachers classifying their own material at the moment they used the tool. What made it break, and what would have made it more likely to hold?

3. What did she buy by recommending the weaker tool to most of the staff, and what did she give up?

STOP HERE

Write your answers before you turn the page. How it played out starts on p. 44.

CASE STUDY · MODULE 1

How it played out

She refused the single answer and split it by content rather than by tool. One sheet, pinned in the staffroom: nothing with a child in it — no names, no scripts, no photographs of work — goes into WhatsApp; everything with no child in it goes wherever you like. She set up the browser tool with training switched off for the nine teachers who wanted it, and taught the setting rather than the product. By December twenty-nine teachers were using the WhatsApp assistant weekly and the nine had accounts. The rule held for a term and then broke in one department, where a teacher photographed a class set to get help marking it; Mrs Verma found out because a colleague mentioned it in passing, not because anything checked. The wrong founding year stayed on the Class 8 wall for three weeks, until a student's grandfather corrected it. She later said that was the most useful thing that happened all term, because she then taught the reliability lesson with the wall as the example and nobody in that class forgot it.

The questions, discussed

1. Both tools gave a founding year, the years differed, and both sounded equally certain. Explain why, from the mechanism.

A founding year for one small town's school has been written down rarely and inconsistently. The model is producing the most likely next words after a phrase like "the school was founded in", and the most likely next words are a plausible year. Nothing in the tone changes between a fact written down a million times and one written down three times, because there is no internal measurement of uncertainty reaching the sentence. That the two tools disagreed was the useful signal: disagreement is a red flag you can trust, while agreement would only have told her the pattern was stable, not true.

2. Her rule depended on teachers classifying their own material at the moment they used the tool. What made it break, and what would have made it more likely to hold?

It broke because it lived on a sheet that nobody re-read at the point of use, and because the tempting case — a stack of scripts and no time — is exactly the case the rule forbids. Rules that survive are the ones attached to the routine rather than to memory: a standing line in the marking procedure, the phone photographs taken with names covered as a matter of course, and someone whose job it is to ask about it once a term. Nothing in the school checked, so the only way anyone learned of the breach was conversation.

3. What did she buy by recommending the weaker tool to most of the staff, and what did she give up?

She bought adoption, and with it the actual benefit — thirty teachers generating misconception lists and worksheets instead of twelve. She also kept the staffroom from splitting into those with accounts and those without. What she gave up was any control over student data in the tool most people were using, and any honest ability to say the training setting was off. She traded a data position she could not enforce for a usage position she could see, and made the trade explicit in writing rather than pretending the two tools were equivalent.

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. 314. If two go wrong, re-read the module before the exam.

- 1 You need three things for a Class 9 worksheet: an explanation of photosynthesis, the year your town's first secondary school opened, and 15 per cent of 4,860. You ask one chatbot, with no web search. Which of the three is safest to use as it comes back?
 - A. The year, because a founding date is a single simple fact and the model either has it or says it does not.
 - B. The explanation, because core school science has been written down consistently by very many people.
 - C. The percentage, because arithmetic follows a rule the model applies rather than a pattern it predicts.
 - D. All three equally, since the model states each one in the same tone and has no reason to be more sure about any of them.

- 2 An answer arrives with three links and a line saying the tool searched the web. What has that changed?
 - A. It has removed the training cutoff and also removed the possibility of an invented citation, because every claim now comes from a page that was actually fetched.
 - B. It has slowed the answer down without changing its accuracy, because the model writes from its training either way and the links are added afterwards.
 - C. It has fixed staleness and given you pages to open; it has not fixed the judgement about which page to trust.
 - D. It has made the answer verified, because a searched answer only contains what the fetched pages say.

- 3 You paste five past papers and ask for ten questions in the style of paper four. Back come ten confident questions in the style of paper one. What most likely happened?
- A. The material did not all fit, so the tool kept part of it and answered as though it had the lot.
 - B. The model prefers the beginning of any long input and was designed to summarise from the first document it is given.
 - C. The free tier cannot read more than one document at a time, and it would have refused had you asked it to.
 - D. Papers two to five were rejected by the safety filter because past papers are copyrighted material, which the model will not reproduce.
- 4 Asked for a Class 10 maths revision sheet, the model produces US Algebra II content, dollars, and the word "sophomore". What is the quickest thing that actually fixes it?
- A. Tell it to be accurate and to follow the Indian curriculum, restating the instruction whenever it drifts.
 - B. Switch to a different tool, since the American default belongs to one company's model rather than to the training text.
 - C. Ask it to search the web for your board's syllabus, because a specific query will bring the board's own document up first and the model will then follow it.
 - D. Paste the board's chapter list and one past paper, and name the terms and units to use.

- 5 You want twenty word problems with a trustworthy answer key. Which approach gives you one?
- A. Ask the model to check its own arithmetic at the end, which catches most carry errors before you see them.
 - B. Use a reasoning mode, which works through the problem in steps and therefore produces exact arithmetic rather than a prediction.
 - C. Ask for the problems and the key in one message, so that the numbers stay consistent with each other throughout.
 - D. Ask for problems whose answers you specify, or work the key yourself with a calculator.
- 6 You ask the same question in three fresh chats and get the same answer three times. What have you established?
- A. That the pattern is peaked, which is not the same thing as the answer being true.
 - B. That the answer is almost certainly right, because three independent runs converged on it.
 - C. That the tool has cached the answer, which is what happens to any question asked more than once.
 - D. That the question is on reliable ground, since agreement across three draws is the same evidence as agreement between two different companies' models.

Figure 2.1

The five things the model cannot reach

Who

Year, subject, class size, language background, reading level.

What constrains the room

Minutes, equipment, layout, whether there is a printer.

What came before

Yesterday's lesson. Tells it where to start and where not to.

The live misconception

The wrong idea you saw in their work, in their words. The most valuable line you can write.

The shape of the output

Exactly what, exactly how many, exactly how long, and what you do not want.

None of these five is in the training text, and all five change the output. The fourth is the one no lesson-plan website contains. Keep it near the top or the bottom of the prompt, because material buried in the middle of a long input is used least.

Class: Form 2, English, 40 students, Kiswahili first language, reading age around 10.
Room: 40 minutes, no projector, I can photocopy one page.
Before this: They wrote a paragraph describing their journey to school. Most paragraphs were a list of events with "and then".
Misconception: They think a good paragraph is a long one. Nobody used a topic sentence.
Give me: Two example paragraphs about the same journey, one that is a list and one with a topic sentence and three supporting details, short enough for a reading age of 10. Then four questions that make students notice the difference.
Do not: Explain what a topic sentence is. I will do that.

That prompt takes ninety seconds to write and produces something you can photocopy. The generic version takes ten seconds and produces something you cannot.

Let it interview you

If you cannot think what to put, reverse the direction. "Before you write anything, ask me five questions about my class and my room that would change what you produce." The model asks; you answer in a line each; then it writes. This works because the model has read a great deal about what makes lessons fail, and it will ask about the things it needs. It also has a limit: left unsteered it asks generic questions — "what are your learning objectives?" — so add "ask about constraints I might not have mentioned, not about objectives".

What to leave out

Do not describe your teaching philosophy. Do not explain that you want the activity to be engaging. Do not write "please" and "thank you" — they cost nothing but they add nothing. Every sentence that is not one of the five things dilutes the ones that are, and the memory-limit lesson explained why: material in the middle of a long input is used less than material at the edges. Keep the misconception near the top or the bottom.

The honest limit

Five things make the output fit your room. They do not make it correct. A perfectly targeted starter can still contain a wrong fraction comparison, and the checking habits from the first block apply to every output, however well-prompted. Prompting well reduces the editing. It does not remove the reading.

EXAMINATION

AI for Teachers: final examination

This paper covers the whole course:

- how a language model works and what follows from that for reliability, arithmetic and memory
- prompting as a teacher
- making explanations, worksheets, question banks and unit plans
- assessment, feedback and marking
- integrity, policy and AI-written work
- teaching AI itself to students from eight to eighteen
- the multilingual, mixed-ability classroom

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 356.

1 Module 1 · Understanding the machine you are about to use

A student says the chatbot lied to her. What is the accurate correction?

- A. It produced a sentence that fits, and a sentence can fit beautifully and be false. It has no way to check.
- B. It did lie, and the companies are working on making the models more honest about what they do not know.
- C. It repeated something false that it read, so the untruth belongs to whoever wrote that page rather than to the machine.
- D. It was misled by whoever asked it the same question before her, because it learns from every conversation it has.

2 Module 1 · Understanding the machine you are about to use

Which one-sentence explanation survives a class asking "then why is it wrong sometimes?"

- A. It is a very large store of facts that has gaps in the places where nobody wrote anything down.
- B. It is a machine that has read an enormous amount of writing by people and guesses what words come next.
- C. It is a robot brain that thinks in a different way from the way we think.
- D. It is like a search engine that finds the best answer available on the internet and puts it in its own words.

3 Module 1 · Understanding the machine you are about to use

A department of six teachers decides to share one free account to save trouble. What happens?

- A. Nothing much, because the message cap is applied per device rather than per account.
- B. The tool suspends the account, because use from several places at once looks like automated activity to the provider.
- C. The answers improve, because the model picks up the department's house style from the shared history.
- D. They hit the message cap by lunchtime, because the cap is per account. Each teacher needs their own.

4 Module 1 · Understanding the machine you are about to use

You switch off the setting labelled something like "improve the model for everyone". What have you achieved?

- A. The tool now runs a model on the phone itself, so nothing you type leaves the device at all.
- B. Your conversations are deleted at the end of each session, which is why the history no longer appears.
- C. What you type is taken out of the training pipeline; the company still receives and stores it.
- D. The conversation is now private, in the sense that nobody at the company is able to read what you send.

5 Module 1 · Understanding the machine you are about to use

A tool that was sharp at seven in the morning is producing thin, shallow work at nine in the evening. What is the usual reason?

- A. Your prompts have got worse as you have got tired, which is the commonest cause of a bad evening session.
- B. Demand is high, so the service is quietly giving you a smaller, faster model.
- C. The model was updated during the day and the newer version is weaker at this kind of task.
- D. You are near the message cap, so the tool is shortening its answers to make your remaining quota last longer.

6 Module 1 · Understanding the machine you are about to use

You are reading a generated answer before using it with a class. What should you mark as unverified?

- A. The passages written in a more confident tone, since hedged sentences mark the places the model is unsure of itself.
- B. Every proper noun, number, date and quotation.
- C. The first and last paragraphs, where errors are known to cluster.
- D. Every sentence containing an absolute word such as "always" or "never".

Module 1 • Understanding the machine you are about to use

The module starts on p. 9.

MODULE 1 TEST

Module 1 test, Q1, p. 46

You need three things for a Class 9 worksheet: an explanation of photosynthesis, the year your town's first secondary school opened, and 15 per cent...

Answer B: The explanation, because core school science has been written down consistently by very many people.

Why: Reliability tracks how many times a thing has been written down and how consistently. Photosynthesis is in every textbook in every language, saying the same thing, so the model's guess is effectively a fact. The founding year of one town's school has been written down rarely, so the model produces a plausible-looking year. And multiplication is not a rule the model applies; it predicts the digits, which is why it slips as the numbers grow. The identical tone across all three is exactly why you cannot use the answer to judge the answer — you judge the question.

Module 1 test, Q2, p. 46

An answer arrives with three links and a line saying the tool searched the web. What has that changed?

Answer C: It has fixed staleness and given you pages to open; it has not fixed the judgement about which page to trust.

Why: Search is a genuine improvement for anything recent or specific, and the real gain is that you now have a page you can open. What it does not change is how the pages were chosen or read: the model ranks and blends them by pattern, so a content farm that copied an old syllabus looks like the exam board, and where three pages disagree the summary may say a fourth thing none of them said. Summaries also lose the exceptions and conditions first, because those are the least predictable words in a sentence.

Module 1 test, Q3, p. 47

You paste five past papers and ask for ten questions in the style of paper four. Back come ten confident questions in the style of...

Answer A: The material did not all fit, so the tool kept part of it and answered as though it had the lot.

Why: Refusal is the good outcome; quiet truncation is the bad one. Everything the model can consider has to fit one window at once, and when it does not, the tool keeps part and answers as though nothing were missing. Nothing warns you. The cheap defence is to give it less and then check what it has: before you start, ask it to tell you the heading of the last section it can see. If it names the wrong section, the material did not fit.

Module 1 test, Q4, p. 47

Asked for a Class 10 maths revision sheet, the model produces US Algebra II content, dollars, and the word "sophomore". What is the quickest thing...

Answer D: Paste the board's chapter list and one past paper, and name the terms and units to use.

Why: The lopsidedness is in the training text, not in one product, so changing tools moves you a small distance and no further. The model is poor at guessing your syllabus and very good at following one it is shown, which is why twenty lines of chapter list and one past paper shift the output almost entirely. Examples beat descriptions here as everywhere: she did not have to explain the Nigerian or Indian school system, only show the shape.

Module 1 test, Q5, p. 48

You want twenty word problems with a trustworthy answer key. Which approach gives you one?

Answer D: Ask for problems whose answers you specify, or work the key yourself with a calculator.

Why: The model is good at the method — setting up the equation, choosing the formula, explaining why — and unreliable at the digits, because arithmetic is a specific that appears rarely in text. Fitting a story to a number you chose is far safer than fitting a number to a story. Reasoning modes do improve arithmetic markedly, but better is not exact, and you cannot see from the outside whether a given number came from a computation or a guess. Asking it to check its own working leaves the same predictor doing both jobs.

Module 1 test, Q6, p. 48

You ask the same question in three fresh chats and get the same answer three times. What have you established?

Answer A: That the pattern is peaked, which is not the same thing as the answer being true.

Why: Disagreement is a red flag you can trust; agreement is a green flag you cannot. Three matching answers tell you the model's learned distribution is sharply peaked, and popular misconceptions are the most stable outputs it has — the ten-per-cent-of-the-brain claim comes back identically every time. Three draws from one system are one source however many times you read it, which is why asking a second company's model is a stronger check than asking the same one again.

THE LESSON CHECKS**Lesson 1, p. 13**

A student tells the AI it made a mistake. It apologises and changes its answer. She says this proves it understood its error. What should...

Answer A: An apology is simply the kind of sentence that follows a correction in the writing it learned from — it will produce one whether or not it was wrong. Tell it a correct answer is wrong and watch it fold.

Why: It predicts likely next words from human writing, so sounding right and being right come apart.

B She is right — noticing and...: Treats fluent self-correction as evidence of an internal grasp of truth, when the same behaviour appears with nothing to correct.

C The company that built it wrote...: Assumes the behaviour comes from hand-written rules rather than from patterns in the text it learned from.

D When challenged, it went and looked...: Imagines the system consults a reference at the moment of asking, like a search engine, rather than generating text.

Lesson 2, p. 16

A teacher on a shared Android phone with patchy data wants to generate practice questions each evening. Which single setup decision matters most before she...

Answer B: Turning off the setting that lets the company train on her conversations, and working in the browser.

Why: Every major model has a free tier that runs in a phone browser or inside WhatsApp; the choice that matters is not which is cleverest but which one you have turned off training on your data.

A Picking whichever tool scored highest on...: Treats capability as the deciding factor; the free tiers of the major tools are close enough for worksheet generation that the difference is swamped by how she prompts and checks.

C Installing every app on the phone...: Assumes storage, data and time she does not have, and comparison across tools is a verification technique for specific doubtful answers, not a daily workflow.

D Paying for one subscription first, since...: Free tiers are the same models with usage caps, not cut-down demos; a teacher generating a few worksheets a night rarely hits the cap.

Lesson 3, p. 21

A geography teacher asks a chatbot two questions: the boiling point of water at sea level, and the population of her town at the last...

Answer C: The boiling point is almost certainly right; the population may be a plausible invention and needs checking against the census.

Why: Reliability tracks how many times a fact has been written down by people — so the model is sound on photosynthesis and unsound on the founding date of your district, and you can predict which before you ask.

A Both are factual questions with definite...: Treats correctness as a property of the question type rather than of how often the answer appears in training text; a rare local figure has been written down a handful of times, a physical constant millions.

B Neither answer can be trusted, because...: Overshoots: a guess informed by millions of consistent examples is reliable in practice, and refusing to distinguish leaves the teacher unable to use the tool at all.

D The population is the more reliable...: Inverts the mechanism; specific numbers with few occurrences in text are among the least reliable outputs, and the model has no arithmetic or lookup behind them.

USE IT IN YOUR WORK

AI for Teachers

For teachers who now have to teach and use AI, often with no training and no projector.

WHAT IT TEACHES	For schoolteachers anywhere who now have to teach and use AI, usually without training.
WHAT IS INSIDE	Seven modules and 57 lessons, 26 figures drawn from data, seven case studies, seven 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-for-teachers

Addaly

First edition, September 2026 · build 62a26075



Scan to open the course