USE IT IN YOUR WORK

AI for Students

How to use these tools and still end up knowing things.

First edition, September 2026

Addaly

Series editor: Dr Mustafa Mun

Draft, open for academic review

USE IT IN YOUR WORK

AI for Students

How to use these tools and still end up knowing things.

Series editor: Dr Mustafa Mun

Addaly

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

AI for Students

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 Students. Addaly. addaly.com/learn/ai-for-students.*

This book is the printed form of the course of the same name at addaly.com/learn/ai-for-students. 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 · 28 figures · 8 case studies · 56,014 words · about 10 hours

Brief contents

1

MODULE 1

What the machine is doing when it helps you

Explain what a language model computes when it answers you, predict from that mechanism which study tasks it will help with and which it will quietly damage, and pick the right free tool for a given piece of schoolwork instead of defaulting to a chatbot.

Lesson 1 · 7 min

Two Kinds of Help

THE ONE THING TO KEEP

After a study session, ask what changed: the document, or you.

The question that actually decides it

Most advice about AI and study starts with rules. Allowed, not allowed, disclose it, don't. That is a real question, and we cover it later. But it is the wrong *first* question, because the answer tells you nothing about what happened to you.

Here is a better one.

After the session, what changed — the document, or you?

Both kinds of help produce a finished assignment. Only one produces a person who could do it again.

The same tool, two directions

Same subject. Same chatbot. Same ten minutes.

Avoiding the work	Doing the work
"Solve question 4."	"I got x = 3, the answer key says x = −1. Here is my working. Which line is wrong? Don't tell me why yet."
"Write an introduction on the causes of the Nigerian Civil War."	"Here is my introduction. Argue against my thesis as harshly as an examiner would."
"Explain recursion."	"Explain recursion, then give me three problems and withhold the solutions."
"Summarise chapter 6."	"I've read chapter 6. Ask me six questions on it, one at a time, and don't accept a vague answer."

The right column is not slower because it is more virtuous. It is slower because you are the one doing the expensive part, and the expensive part is the point.

The load-bearing part

Every assignment has one thing it exists to build. Find it, and you know what not to hand over.

- In a problem set, the load-bearing part is the solving. Not the arithmetic, not the neat presentation.
- In an essay, it is the argument — which claims you make, in what order, and what you leave out.
- In a lab report, it is the interpretation. Not the method section you have written eleven times.
- In a translation exercise, it is choosing between two defensible renderings.

Everything else is scaffolding, and scaffolding is fair game. Reformatting a bibliography into APA is not a skill anybody is examining. Turning your bullet points into a table is not either. If you spend the saved hour on the load-bearing part, the tool has done its job.

The transfer test

Here is a check you can run today, and it costs ten minutes.

Finish a piece of work with AI help. Close the laptop. Wait ten minutes — get water, walk somewhere. Then take blank paper and redo the central thing: solve the problem again, or write your thesis and its three supports from memory.

Whatever comes back is yours. Whatever does not, you watched somebody else do.

This is uncomfortable the first time, because the gap is usually much bigger than it felt. Reading a clear explanation produces a strong, immediate feeling of understanding. That feeling is real, and it is not knowledge. Cognitive

psychologists call it the fluency illusion, and AI is very good at generating it, because AI is very good at being clear.

Neither extreme survives contact with reality

Refusing to touch these tools is not integrity, it is just a tax you pay in hours. A student in Lagos or Lima or Lucknow with no tutor, no office hours, and a class of 300 has something now that did not exist before: a patient thing that will explain the same idea six different ways at two in the morning. Telling that student to switch it off is not advice, it is a lecture.

And going the other way — running every task through the model and shipping the output — reliably produces someone who cannot tell when the output is wrong. That is not a moral failure. It is a mechanical one, and we will look at the evidence in the next lesson.

The useful position is in the middle and it is specific: **delegate the scaffolding, keep the load-bearing part, and test yourself on paper afterwards.**

One habit to start with

Before you type anything into a chat box, finish this sentence out loud:

> *"I am asking for this because I want to be able to _____ afterwards."*

If the honest ending is "submit it," you already know what kind of help this is. That is not always the wrong answer — some weeks are like that. But say it plainly to yourself, because the cost is real and it arrives later, in a room where the tool is not allowed.

BEFORE YOU MOVE ON

Priya asks a chatbot to solve five calculus problems and reads each solution until every step makes sense. Ravi solves the same five himself, gets three wrong, then asks only what went wrong with his method. A week later there is an unassisted test. Why is Ravi likely to do better?

A. Priya's mistake was not putting the solutions into her own words, so none of it became hers.

B. Ravi got three wrong, and mistakes are what teach; had he solved all five correctly he would have gained nothing.

C. Ravi produced the steps himself, and producing something under difficulty is what makes it available later; Priya recognised steps she never generated.

D. Priya did not read deeply enough. Had she studied each solution until it was completely clear, she would be equally prepared.

The answer and why: p. 393

Lesson 2 · 8 min

What Happens To Your Memory

THE ONE THING TO KEEP

The tutor that gave answers raised practice scores and lowered exam scores; the one that gave hints did neither harm.

The study you should know about

In 2024, researchers ran something close to the experiment every student privately wonders about. Bastani and colleagues worked with roughly a thousand high school students in Turkey, across grades 9 to 11, in maths.

Students were split three ways for their practice sessions:

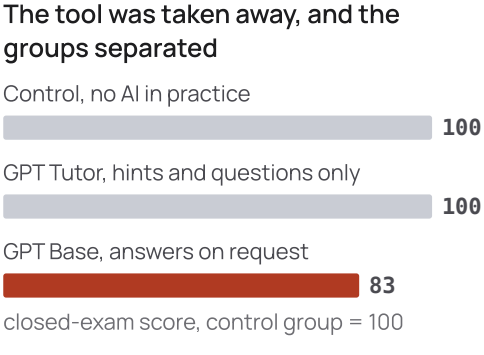
- **Control** — practice as normal, no AI.
- **GPT Base** — a standard GPT-4 chat assistant. Ask anything, get an answer.
- **GPT Tutor** — the same underlying model, but constrained: it gave hints, asked questions, and would not hand over the solution.

During practice, both AI groups did better than the control group. Not slightly better. GPT Base students got about **48% more practice problems right**. GPT Tutor students got about **127% more right**. If you stopped the study there, you would write a press release.

Then came an exam. Closed. No AI for anyone.

The GPT Base students scored about **17% worse than students who had never had the tool at all**. The GPT Tutor students came out roughly level with the control group — the safeguards removed the harm, though they did not produce a lasting gain either.

Figure 1.1



Bastani and colleagues, about a thousand Turkish high-school students, mathematics, 2024, with the control group set to 100. During practice both assisted groups did better, and the gap was not small: the answering assistant raised practice accuracy by roughly half, the hinting one by more than double. Then came a closed exam with no tool for anyone. The variable that moved the exam score was not whether AI was present. It was whether it gave answers or gave hints, and that is the part you set in the prompt.

Read that carefully, because the obvious reading is wrong

The intuitive story is "the AI group was lazy." That is not what happened. The AI group *worked and performed better* during practice. They were not doing less. They were doing something else.

The intuitive story two is "AI is bad for learning." That is not what happened either — the two arms used the same model. The variable that moved the exam score was not whether AI was present. It was **whether it gave answers or gave hints**.

That is a much more useful finding, because it is something you control. You cannot change what model your school has access to. You can change what you ask it for.

Why practice performance and learning came apart

This is not a strange new effect of AI. It is a well-documented feature of how memory works, and it has a name: **desirable difficulties**, from Robert Bjork's work in the 1990s.

The short version: the conditions that make performance *during* study feel good and look good are frequently the conditions that produce the least durable learning. Conditions that feel slow and error-strewn — recalling from memory, spacing sessions out, mixing topics — feel worse and work better.

An AI that answers is a machine for making study feel good. Every step arrives clean. You never sit in the gap. And the sitting in the gap, it turns out, was not an obstacle to the learning. It was the learning.

There is a related result worth knowing: Roediger and Karpicke (2006) had students either reread a passage repeatedly or read it once and then try to recall it. Five minutes later, the rereaders won. A week later, the recall group remembered substantially more — and, notably, the rereaders had *predicted* they would do better. The feeling of readiness pointed the wrong way.

What is not settled

Be honest about the state of the evidence, because you will meet people who overclaim in both directions.

The Turkey study is one study. One subject, one country, one model version, one age group. Maths is unusually well-suited to the design; nobody has shown the same numbers for essay writing or lab work. A 17% drop is large, and single large effects sometimes shrink when repeated.

A widely shared 2025 MIT Media Lab preprint ("Your Brain on ChatGPT") reported weaker EEG connectivity in essay-writers using an LLM, and that most of that group could not quote a sentence from an essay they had finished minutes earlier. It is suggestive. It is also 54 participants and, at release, not peer-reviewed. Do not treat brain scans as proof of anything about your degree.

What is solid is the underlying mechanism, and it has forty years behind it: **you remember what you generate, not what you read.** The Turkey study is a clean demonstration of that principle meeting a new tool.

The practical conclusion

You do not have to give up the tool. You have to change what you ask it to be.

Every time you are about to type a question, you are choosing between GPT Base and GPT Tutor. The model does not choose. You do, in the prompt. The next lesson is about how.

BEFORE YOU MOVE ON

In the Turkey study, unguarded-tutor students solved 48% more practice problems correctly than the control group and then scored 17% worse on the unassisted exam. What does the hint-only arm add to that picture?

A. That practice scores were an unreliable measure, so neither group's practice gain tells you anything useful.

B. That the harm came from how the tool answered rather than from AI use as such — same model, hints instead of solutions, large practice gains, no exam penalty.

C. That the unguarded group simply did less work, and the volume difference explains the exam gap.

D. That the effect belongs to that particular model version, so a more capable model would remove the penalty.

The answer and why: p. 394

Lesson 3 · 9 min

What it is actually doing

THE ONE THING TO KEEP

A language model predicts the next token from a compressed statistical impression of its training text, so it is strong on repeated general material, weak on specific rare facts, and blind to the letters inside words.

One token at a time

A language model does one thing. Given the text so far, it produces a probability for every possible next chunk of text, and one chunk is chosen. Then it does it again, with that chunk added to the text. Then again. An answer that took you fifteen seconds to read was built in a few hundred of these steps, each of which knew nothing about the ones still to come.

CASE STUDY · MODULE 1

The tutorial sheet that everyone finished

An illustrative case, built from documented patterns.

A government engineering college in Tiruchirappalli, where 180 first-year students share one maths tutor and a weekly problem sheet worth fifteen per cent of the module.

In the second week of the semester, Dr Revathi Sundaram noticed that her first-year tutorial sheets had started coming back correct.

That is not, in itself, a complaint. Sheet one had averaged forty-one out of a hundred, which was normal and depressing. Sheet four averaged eighty-nine. The handwriting was still bad and the presentation was still bad, and the answers were right.

She had 180 students, one tutorial slot a week and no teaching assistant. The sheets carried fifteen per cent of the module. The January internal examination was four weeks away, and it is closed, handwritten and invigilated.

A short conversation in the corridor confirmed the obvious explanation. Most of the class was using a free chat tool on a phone, in a hostel room, at night.

Her first instinct was to prohibit it, and she spent an evening working out how. Enforcement was the smaller problem, though it was real enough: she could not invigilate a hostel. The argument that stopped her came from a student. He said that his parents both worked nights, that nobody at home had finished school, and that this was the first time in his life he had been able to ask a question at eleven at night and get an answer.

She had taught for nineteen years in a college where the students who do best are usually the ones with an engineer somewhere in the family. A patient thing that explains the same idea six different ways at two in the morning is, for exactly her students, the most levelling development of her career. She was not going to switch it off and call that integrity.

So the decision was never ban or allow. It was what the sheet is for.

The first option was to change nothing and accept that fifteen per cent of the module now measured something other than what it used to. That is cheap and it produces a cohort whose coursework marks and examination marks disagree violently in January, which is a problem that arrives too late to do anything about.

The second was to keep the sheets, take the marks off them, and move the fifteen per cent onto a ten-minute handwritten problem at the start of every tutorial: closed, cold, phones in bags. That cost her 180 short scripts to mark every week for the rest of the semester, on top of everything else she was carrying. It cost the students something sharper. Several were relying on sheet marks as a cushion, and taking a cushion away in December, from students whose families are paying for the year, is not a neutral administrative act.

One further thing decided her, and she found it while reading sheet four rather than while thinking about policy. A substantial number of the correct sheets carried one wrong number inside otherwise correct working. The setup was right, the method was right, the final arithmetic was out by a digit or a factor. Nobody had checked. Her students had learned to trust the machine at precisely the task it is worst at, and they could not have told you that, because nothing in the output looks different when it is wrong.

She put the second option to the class herself, in the tutorial, with the sheet-four statistics written on the board, and told them why.

YOUR ANSWERS

1. The same students in the same week scored eighty-nine assisted and thirty-four cold. What does the difference measure, and why could neither number on its own have told her anything useful?

2. Many correct sheets carried one wrong number inside correct working. Explain from the mechanism why arithmetic is where that happens, and name what a student should have done instead.

3. Dr Sundaram rejected a ban on a fairness argument rather than an enforcement one. Reconstruct that argument, and say what it does and does not license.

STOP HERE

Write your answers before you turn the page. How it played out starts on p. 50.

CASE STUDY · MODULE 1

How it played out

She ran it from week five. In that first week the ten-minute cold problems averaged thirty-four, against a sheet average of eighty-nine from the same students in the same week. By week twelve the cold average was fifty-eight and the sheet average had barely moved. The gap did not close because the sheets got worse. It closed because students began, without being instructed to, doing each sheet twice: once with help and once on blank paper the next day.

Eleven students held sheet marks above ninety and cold marks below twenty-five for three consecutive weeks. She called them in one at a time and accused nobody of anything, and in nine of the eleven conversations the student said a version of the same sentence: they had understood it at the time. The internal examination averaged fifty-one against forty-seven the previous year, which she declines to present as evidence of anything, being one cohort and one year.

Two things she would change. She gave the class a standing instruction to paste into the tool in week five — hints and questions only, full solutions only when I type SOLVE — and it was widely ignored, because an instruction nobody yet has a reason to want is a suggestion. It started being used in week nine, once the cold problems had made the reason obvious, so she now introduces the cold problems first and the instruction second. And two students stopped attending tutorials rather than sit a cold problem every week. She considers that a cost she has not solved.

The questions, discussed

1. The same students in the same week scored eighty-nine assisted and thirty-four cold. What does the difference measure, and why could neither number on its own have told her anything useful?

The eighty-nine measures performance under the conditions of practice, with a fluent explainer available. The thirty-four measures capability under the conditions of the assessment. They are different quantities and they come apart, which is the finding from the Turkey study: an assistant that answers

raises practice accuracy sharply and lowers closed-exam performance, while one that hints does neither harm nor lasting good. The assisted number on its own is an advertisement, because every incentive in the moment makes assisted work feel like learning. The cold number on its own is a grade with no diagnosis attached. It is the gap between them, measured on the same students in the same week, that tells her which students are in trouble and roughly how much.

2. Many correct sheets carried one wrong number inside correct working. Explain from the mechanism why arithmetic is where that happens, and name what a student should have done instead.

Numbers are split into tokens that do not respect place value, so the model never sees a column of digits to carry between. It produces the most plausible continuation, which for small familiar numbers is almost always right and degrades as the digits multiply. Nothing in the tone changes when it degrades, so a wrong number sits in the middle of correct working looking exactly like a right one. The fix is a division of labour: take the method, the setup and the interpretation from the model, and take the number from something that computes rather than predicts — SymPy or Maxima, both free, or a code interpreter that writes Python and runs it. With no computer at all, three checks cost nothing: estimate the order of magnitude first, carry the units down every line, and substitute the answer back into the original equation.

3. Dr Sundaram rejected a ban on a fairness argument rather than an enforcement one. Reconstruct that argument, and say what it does and does not license.

The argument is that the students who lose most from a prohibition are the ones with no alternative source of help: no tutor, no graduate at home, no office hours that are genuinely available. For them the tool is not a shortcut around teaching they could otherwise get, it is the only teaching available at eleven at night, and removing it widens exactly the gap the college exists to narrow. That licenses refusing a ban, and it licenses treating access as a good. It does not license leaving the assessment untouched, because the fairness argument is about access to explanation, not about what a mark should

measure. Her response keeps the access and moves the marks onto something the tool cannot sit inside, which is the only combination that answers both halves.

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

1 In the Turkey study, both AI groups outperformed the control group during practice and then diverged on a closed exam. What was the variable that moved the exam score?

A. Whether the assistant gave answers or gave hints
B. Whether students had an assistant during practice at all
C. How many practice problems each group answered correctly
D. How much total time each group spent in practice sessions

2 A model that can discuss protein folding miscounts the letter r in strawberry. Why?

A. Counting is arithmetic, and it has no calculator attached
B. The word is too rare to have been in its training data
C. It sees tokens, not the letters inside them
D. Its temperature setting makes each answer different

3 Fabricated references are far more common than fabricated general explanations. Why does the risk concentrate there?

A. Reference databases were left out of the training data on purpose
B. A citation is a rigid form whose specifics are rare
C. The model is protecting the copyright of papers it was trained on
D. Reference lists come last, where the model has run low on context

4 You upload a forty-page PDF and get poor answers about page twenty-two. What is the fix?

A. Upload the whole document again and ask one global question
B. Ask the model to read the document twice before answering
C. Raise the temperature so it samples more of the context
D. Paste the two or three pages that contain the answer

5 Which single question sorts most of schoolwork into the right tool without needing a list?

A. What would a confidently wrong answer cost me here?
B. Is this the sort of task a chatbot is generally good at?
C. Has the model been updated since I last asked this?
D. Could I explain this answer to somebody afterwards?

6 Why is asking 'are you sure?' not a check on an answer?

A. It consults a stored confidence score and reports it honestly
B. It re-runs the computation and reports whether the two agree
C. It is another prompt, and agreeableness invites a retraction
D. It quietly searches the web before it answers the challenge

Lesson 10 · 9 min

The anatomy of a study prompt

THE ONE THING TO KEEP

State level, task, constraints and form, include your current wrong understanding, and when an answer disappoints work out which of those four you left out.

Four things the model cannot guess

A weak prompt is not usually too short. It is missing information the model has no way to infer and will therefore invent: who you are, what you already know, what the answer is for, and what shape it should take.

Compare.

> *Explain enzyme inhibition.*

against

> *I am in the second year of a biology degree. I understand Michaelis-Menten kinetics and can read a Lineweaver-Burk plot. Explain the difference between competitive and non-competitive inhibition in terms of what happens to Vmax and Km, and why. Around 250 words. Do not define enzyme or substrate. End with one question that would catch me if I had only memorised the result rather than understood it.*

The first produces a textbook page you have already read. The second produces something you can use. Nothing in it is clever. It states four things:

Figure 2.1

Four things the model cannot guess

Explain enzyme inhibition.

- No level, so it starts at the beginning
- No task, so it tours the topic instead of answering a question
- No constraints, so it runs to a page and defines what you know
- No form, so it arrives as a balanced mid-length essay
- You get the textbook page you have already read

The same question with the four supplied

- Level: second year, reads a Lineweaver-Burk plot
- Task: what happens to Vmax and Km, and why
- Constraints: 250 words, do not define substrate
- Form: prose, ending in one question that would catch me
- Plus the wrong version you currently hold, so it diagnoses

Nothing in the right column is clever, and prompt writing is not a craft with rituals. What it is, is a habit of noticing which of the four you left out. An answer that is too basic means level. One that wanders means task. One that is too long means constraints. One in the wrong shape means form. Diagnosing takes five seconds and beats rewording at random.

1. **Level and prior knowledge** — so the answer starts where you are, not at the beginning.
2. **The specific task** — not the topic, the job. *Explain the difference in terms of Vmax and Km.*
3. **Constraints** — length, what to omit, what to assume.

EXAMINATION

AI for Students — final examination

This paper covers the whole course:

- what a language model computes and which study tasks that makes it good and bad at
- how to write a study prompt and ground it in a source you supply
- how to run a session in which the assistant raises the difficulty instead of lowering it
- the characteristic failure in each subject and the check that catches it
- taking a claim to a source you have opened
- writing whose argument and structure are yours
- your institution's rules and the record that protects you
- preparing for assessment that happens with nothing to consult

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 441.

1 Module 1 · What the machine is doing when it helps you

In a lab report, which part is load-bearing — the part the assignment exists to build?

A. The interpretation of the results

B. The method section you have written eleven times

C. Reformatting the bibliography into APA style

D. The presentation of the results table

2 Module 1 · What the machine is doing when it helps you

A model explains something clearly and you feel you have understood it. What is the only honest test?

A. Rate your understanding out of seven before and after

B. Close the window and write the explanation from memory

C. Ask the model three comprehension questions about it

D. Re-read the explanation once more the following morning

3 Module 1 · What the machine is doing when it helps you

Why does the same free tier feel less generous in Hindi, Bengali or Tamil than in English?

A. Providers throttle those languages to manage demand

B. The model translates internally, which doubles the work

C. Non-English answers are routed to a slower model tier

D. The same meaning costs more tokens in those scripts

4 Module 1 · What the machine is doing when it helps you

You run a small model locally on an eight-gigabyte laptop. Which job is it genuinely adequate for?

A. Quizzing you from a chapter you paste in

B. Answering questions that need broad world knowledge

C. Following long multi-step instructions reliably

D. Summarising a sixty-page document in one pass

5 Module 1 · What the machine is doing when it helps you

What is the test for whether a capability belongs on your never-outsource list?

A. Whether the skill appears in the module's learning outcomes

B. Whether a model currently does it worse than you do

C. Whether you could tell if somebody else did it badly

D. Whether the department examines it in a closed paper

6 Module 1 · What the machine is doing when it helps you

You ask the same factual question twice, cold, in two fresh chats, and get two different answers. What have you learned?

A. Two agreeing answers would have proved the claim established

B. Two disagreeing answers mean you need a real source

C. Two disagreeing answers mean the temperature is too high

D. Asking twice in the same chat would have worked as well

7 Module 1 · What the machine is doing when it helps you

What is the ten-second habit to apply to any answer you are about to rely on?

A. Ask the model which part it is least confident about

B. Check whether the answer is longer than you expected

C. Underline every proper noun, number and reference

D. Re-read your prompt to see whether it was ambiguous

8 Module 2 · Asking well

You ask for the same idea below your level, at your level, and above it. What does the third pass give you that the others do not?

A. A simpler analogy that a younger student would follow

B. The standard treatment in your course's own notation

C. A list of the specialist papers you would have to read

D. Which parts of what you learned are approximations

Module 1 · What the machine is doing when it helps you

The module starts on p. 11.

MODULE 1 TEST

Module 1 test, Q1, p. 53

In the Turkey study, both AI groups outperformed the control group during practice and then diverged on a closed exam. What was the variable that…

Answer A: Whether the assistant gave answers or gave hints

Why: Both AI arms used the same underlying model, so the presence of AI was not the variable. GPT Base students scored about seventeen per cent below students who had never had the tool; GPT Tutor students came out roughly level with the control group. The difference between the arms was the constraint — hints and questions rather than solutions — which is the part you set yourself, in the prompt, every time you type one.

Module 1 test, Q2, p. 53

A model that can discuss protein folding miscounts the letter r in strawberry. Why?

Answer C: It sees tokens, not the letters inside them

Why: The model never saw the characters. Strawberry arrives as something like str, aw, berry, so asking it to count letters is like asking you to count the pen strokes in a word somebody read aloud to you. The same blindness affects counting the words in your essay, alphabetising a list, and anything that turns on spelling, so use something that operates on characters: a word counter, a spreadsheet, or a code tool that writes and runs a program.

Module 1 test, Q3, p. 53

Fabricated references are far more common than fabricated general explanations. Why does the risk concentrate there?

Answer B: A citation is a rigid form whose specifics are rare

Why: Every component is drawn from a strong pattern: authors who publish in that area, a year in the plausible range, a title built from the field's vocabulary, a journal that genuinely publishes that subject, a DOI prefix that is real. Rigid forms are what a next-token predictor reproduces most easily, while the particular combination was stated once or never. There is no citation feature and no lookup — it is the process that writes the prose, applied to a highly structured string.

Module 1 test, Q4, p. 54

You upload a forty-page PDF and get poor answers about page twenty-two. What is the fix?

Answer D: Paste the two or three pages that contain the answer

Why: A large window means the text fits, not that it is used evenly. Repeated experiments find that material at the very start and the very end of a long context is used reliably and material buried in the middle much less. The fix is not a bigger window, it is putting less in — which also gives faster replies and stretches a free tier, because quotas are measured in the same tokens the window is.

Module 1 test, Q5, p. 54

Which single question sorts most of schoolwork into the right tool without needing a list?

Answer A: What would a confidently wrong answer cost me here?

Why: If the honest answer is a wasted five minutes, use the chatbot and move on. If it is a wrong number in a lab report, a fabricated source in a bibliography, or a program that runs and produces rubbish, use the deterministic tool — SymPy, jamovi, the DOI resolver — and let the model advise you about which one. The last option is the transfer test from the first lesson, which is a good question about learning and does not tell you which instrument to pick up.

Module 1 test, Q6, p. 54

Why is asking 'are you sure?' not a check on an answer?

Answer C: It is another prompt, and agreeableness invites a retraction

Why: There is no lookup to consult and no confidence dial connected to the output text. A challenge is simply more context, and the most plausible continuation of a challenge is frequently an apology and a different answer, whether or not the first one was wrong — you can talk a correct model out of a correct answer, which tells you how much information the retraction carries. What works instead is asking for something checkable, or asking again cold in a fresh chat.

THE LESSON CHECKS

Lesson 1, p. 15

Priya asks a chatbot to solve five calculus problems and reads each solution until every step makes sense. Ravi solves the same five himself, gets…

Answer C: Ravi produced the steps himself, and producing something under difficulty is what makes it available later; Priya recognised steps she never generated.

Why: After a study session, ask what changed: the document, or you.

A Priya's mistake was not putting the…: Treats the problem as one of paraphrasing and ownership, as if learning were a plagiarism question rather than a memory one.

B Ravi got three wrong, and mistakes…: Over-reads the advice to learn from mistakes, making error the mechanism when the mechanism is effortful retrieval, right or wrong.

D Priya did not read deeply enough…: Assumes clarity while reading equals knowing, which is exactly the fluency illusion a well-written solution creates.

Lesson 2, p. 19

In the Turkey study, unguarded-tutor students solved 48% more practice problems correctly than the control group and then scored 17% worse on the unassisted exam…

Answer B: That the harm came from how the tool answered rather than from AI use as such — same model, hints instead of solutions, large practice gains, no exam penalty.

Why: The tutor that gave answers raised practice scores and lowered exam scores; the one that gave hints did neither harm.

A That practice scores were an unreliable…: Reads the result as discrediting the practice measure, when the point is that practice performance and durable learning genuinely came apart.

C That the unguarded group simply did…: Assumes the AI group was lazier, but they performed better during practice, not worse — they were doing something else, not less.

D That the effect belongs to that…: Treats model capability as the variable when the manipulated variable was answer-giving versus hint-giving, which a stronger model does not change.

Lesson 3, p. 22

A student asks a model to check whether their 500-word limit essay is under the limit. It says 478 words. The real count is 611…

Answer D: It reads text as tokens, not words, so it is estimating a plausible number rather than counting.

Why: A language model predicts the next token from a compressed statistical impression of its training text, so it is strong on repeated general material, weak on specific rare facts, and blind to the letters inside words.

A The model was trained on older…: Treats counting as a fact to be remembered rather than an operation to be performed; training date has nothing to do with arithmetic on the text in front of it.

B The essay exceeded the context window…: Plausible-sounding but wrong scale: a 611-word essay is roughly 800 tokens, far inside any modern context window.

C Temperature was set too high, so…: Confuses variability with capability; at temperature 0 the model would give a stable wrong count, not a right one.

USE IT IN YOUR WORK

AI for Students

How to use these tools and still end up knowing things.

WHAT IT TEACHES	A course for school and university students on using AI without losing the ability to think.
WHAT IS INSIDE	Eight modules and 73 lessons, 28 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-for-students**

Scan to open the course

Addaly

First edition, September 2026 · build 62a26075