ADDALY · THE AI CLASSROOM

AI for Students

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

---

8 modules · 73 lessons · 28 figures · 8 case studies · 56,014 words · about 10 hours

Dr Mustafa Mun
Author and editor

ABOUT THIS BOOK

AI for Students, published by Addaly, 2026. Author and editor: **Dr Mustafa Mun**.

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.

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. If you paid for this, somebody has sold you something that was given away.

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. Answers to every exercise, test and examination question are at the back, with a note on why each wrong option attracts people — those notes are the teaching, not the marking.

Contents

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

2. Asking well

Write a study prompt that states the task, the level, the constraints and the required form, ground it in a source you supply rather than the model's memory, and diagnose from a bad answer whether the fault was the model's or your instruction's

1. The anatomy of a study prompt
2. Give it the source
3. Ask for the working, then check it
4. Follow-ups that find the weak joint
5. Explain it at three levels
6. Conversations have a shape
7. What not to paste
8. Diagnosing a bad answer
9. Prompts worth keeping
10. Case study: Nine days, and one thousand nine hundred and fifty rupees
11. Module test

3. Study that raises the difficulty

Run a study session in which the assistant raises the difficulty rather than lowering it — a hint ladder instead of a solution, retrieval before review, spaced and interleaved practice, explanation produced by you — and demonstrate afterwards what you can do with the window closed

1. Getting Unstuck, Not Getting Finished
2. Make It Quiz You
3. Building a hint ladder
4. Retrieval before review
5. Spacing and interleaving
6. Explaining it back
7. Worked examples, and when to stop using them
8. A study session that works
9. Study groups, with and without the machine
10. Proving it to yourself
11. Case study: The evening they stopped finishing things
12. Module test

4. Subject by subject

Name the characteristic failure mode of an AI assistant in a given subject — arithmetic in maths, plausible invalid steps in proof, code that runs and is wrong, invented quotations in history, fossilised errors in language learning — and apply the specific check that catches it, using a free tool where a deterministic one is required

1. Maths, and the arithmetic problem
2. Proofs, and the step that looks fine
3. Code that runs and is wrong
4. The lab sciences, and the dimension check
5. History, and the quotation that was never said
6. Literature, and the text it half remembers
7. Languages, where it is genuinely excellent
8. Statistics for coursework
9. The essay subjects, and what a marker is reading for
10. Case study: Thirty-four participants and a ceiling
11. Module test

5. Research, sources and evidence

Take a claim from a model to a verifiable primary source using free tools, judge what strength of evidence that source carries, and build a bibliography in which you have opened every item

1. Research, And Sources That Exist
2. How a citation gets invented
3. Finding the paper, for nothing
4. Reading a paper without reading all of it
5. What counts as evidence
6. When it searches the web
7. Notes that survive the term
8. Building a bibliography you can defend
9. Misinformation, and what you pass on
10. Case study: Five authorities nobody had opened
11. Module test

6. Writing, and keeping your voice

Produce written work whose argument, structure and sentences are yours, use assistance at the specific points that do not displace your thinking, and show a process record demonstrating how the piece was made

1. Writing With It Without Losing Your Voice
2. Where the thinking happens
3. The blank page, without outsourcing it
4. Structure that carries an argument
5. Editing your own prose
6. Your voice, measurably
7. Feedback that is actually useful
8. Presentations, posters and visuals
9. Group work, and shared rules
10. Case study: The section nobody would claim
11. Module test

7. Rules, honesty and the record

Read your institution's AI policy and place a given piece of work on the permitted, disclosed or prohibited scale, write a disclosure statement, explain from its mechanism why a detector score is not evidence, and keep a process record that would survive an academic-integrity hearing

1. The Rules, And Why Detectors Do Not Work
2. Reading your actual policy
3. Writing a disclosure that helps you
4. How detectors work, and why they cannot
5. The process record
6. If you are accused
7. Where the line really is
8. Accessibility, fairness and who these rules fall on
9. What happens to what you type
10. Case study: Nineteen flags and four days
11. Module test

8. Assessment, and the years after it

Prepare for closed and oral assessment by practising under the conditions of the test, manage a long project with a defensible record, and decide which capabilities to build unaided given what the tools can and cannot do

1. Revision That Survives The Exam Hall
2. The Honest Case For Still Learning It
3. Past papers and mark schemes
4. Practising under the conditions of the test
5. Vivas, and being asked about your own work
6. Long projects and dissertations
7. Applications, statements and interviews
8. Placements, and rules that are not your university's
9. What to learn anyway
10. Case study: Ten minutes each, two hundred and forty students
11. Module test

AI for Students — final examination

The paper that opens the next course. Answers to everything follow it.

MODULE 1

What the machine is doing when it helps you

Before any study technique, the mechanism: what a language model computes, why it sounds certain when it is wrong, what that does to your memory, and which tool actually belongs to which job.

> BY THE END OF THIS MODULE YOU CAN
>
> 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

---

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.

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.

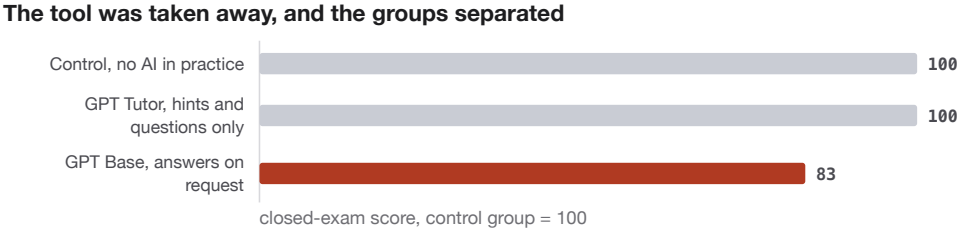
Students were split three ways for their practice sessions:

- **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.



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.

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.

The chunks are called **tokens**. A token is roughly four characters of English, so about three quarters of a word. Common words are single tokens; unusual ones split. `understanding` is usually one token. `Rutherfordium` is four or five. A Hindi or Tamil sentence often costs two to three times as many tokens as the same sentence in English, which is why some tools feel slower and hit limits faster in those languages. Nobody chose that as policy. It falls out of how the vocabulary was built, mostly from English text.

You can see the split for yourself. OpenAI's tokenizer page and the free `tiktoken` Python library both show it; so does Hugging Face's tokenizer playground, which is free and needs no account.

Why it cannot count the letters in a word

Ask a model how many times the letter r appears in *strawberry* and there is a decent chance of a wrong answer, even from a model that will happily discuss protein folding. This is not stupidity. The model never saw the letters. It saw the tokens `str`, `aw`, `berry`. Asking it to count characters is like asking you to count the pen strokes in a word somebody read aloud to you.

This matters more than it looks. The same blindness affects counting words in your essay, alphabetising a list, working with the digits of a long number, and any puzzle that turns on spelling. If you need those, use something that operates on characters: a word counter, a spreadsheet, `wc -w` in a terminal, or the model's code tool if it has one, which writes a Python program and runs it.

Where the knowledge sits

There is no database inside. During training, the model adjusted billions of numerical weights so that its next-token predictions matched an enormous quantity of text. What survives is a compressed statistical impression of that text. Facts repeated in thousands of places — the boiling point of water, the year the Second World War ended — are represented strongly and come out reliably. Facts stated once, in one obscure page, are represented weakly or not at all, and the model has no way to tell the difference between recalling one and constructing something that looks like one.

That is why it is excellent on the material in the first two chapters of your textbook and unreliable on the specific figure from the specific 2019 paper your lecturer mentioned.

Temperature, and why the same question gives different answers

At each step the model has a probability distribution over tokens. **Temperature** controls how much it favours the most likely one. At temperature 0 it takes the top token every time and the same prompt gives nearly the same answer. Higher temperature samples more widely and produces more variety.

Most chat interfaces run somewhere in the middle, which is why asking the same question twice can give you two different answers, and occasionally one right and one wrong. This is worth exploiting: if you are unsure about an answer, open a fresh conversation and ask again in different words. Two independent agreeing answers are weak evidence. Two disagreeing answers are strong evidence that you need a real source.

What it does not have

- **No memory between conversations**, unless a memory feature is switched on. A new chat starts blank.
- **No sense of time.** Its training data stops at some date. It may not know that, and will often answer questions about last month as confidently as questions about last century.
- **No access to your file** unless you paste it or upload it.
- **No internal signal for truth.** It has a signal for *plausible continuation*. Those coincide often enough to be useful and diverge often enough to be dangerous.

The practical consequence

Every reliable technique in this course follows from this one mechanism. You ground it in text you supply, because supplied text is stronger than compressed memory. You ask for working, because a chain you can inspect is

checkable and a conclusion is not. You verify anything specific, because specific things are the weakly-represented ones. You never let it do the counting.

A model is a fluent, widely-read, extremely fast colleague with no memory, no source list, and no ability to tell you when it is guessing. Treat it exactly like that and most of the trouble disappears.

BEFORE YOU MOVE ON

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. What is the underlying reason?

A. The model was trained on older essays and has drifted, so its counting is out of date.

B. The essay exceeded the context window, so the model only saw part of it and counted what it could see.

C. Temperature was set too high, so the model sampled a random number instead of the correct one.

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

4 · 9 min

Why it invents things, and when

THE ONE THING TO KEEP

Fabrication is the model doing its job — producing plausible text — so risk concentrates in names, numbers, dates and references, not in the general explanation around them.

Fabrication is the system working normally

A model that produces a fake reference is not malfunctioning. It is doing exactly what it was trained to do: produce the most plausible continuation. The most plausible continuation of *a citation for the claim that spaced repetition improves retention is* is a citation-shaped string. Author surname, initials, a year, a journal that publishes that sort of work, a plausible volume and page range. Every part is drawn from a well-learned pattern. The combination has never existed.

This is the single most useful thing to understand about these tools, because it tells you exactly where the risk lives. The risk is not in the reasoning. It is in the specifics.

The shape of the risk

Rank these by how likely the model is to be wrong, and the ranking is stable across every model you will use:

1. **Very safe.** General explanations of well-established material. What photosynthesis does. Why a for-loop is slower than vectorised code. The structure of a five-paragraph argument.
2. **Usually safe.** Standard worked methods. How to integrate by parts. How to balance a redox equation. These are procedures repeated thousands of times in the training text.

3. **Risky.** Anything with a number, a name, a date, or a page reference attached. Population of a city in a particular year. Which scholar first argued a position. The exact wording of a statute.
4. **Very risky.** Anything that must exist in the world for your claim to hold — a paper, a URL, a quotation, a court case, a library function, a command-line flag.

Categories 3 and 4 are where the failures live, and they are also the parts a marker checks.

Where the risk sits inside one answer

Very safe: general explanation of well-established material	What photosynthesis does. Why a loop is slower than vectorised code.
Usually safe: standard worked methods	Integration by parts. Balancing a redox equation. Repeated thousands of times in the training text.
Risky: anything with a number, a name, a date or a page attached	A city's population in a given year. Who first argued a position. The wording of a statute.
Very risky: anything that must exist in the world for your claim to hold	A paper, a URL, a quotation, a court case, a library function, a command-line flag.

One paragraph usually contains all four layers at once, written in the same confident register, because there is no confidence dial connected to the output text. Underlining every proper noun, number, date and reference marks the bottom two layers in about ten seconds, and those are the parts a marker checks.

Why it sounds certain

There is no confidence dial connected to the output text. The model produces the tone that fits the context, and academic-sounding questions have academic-sounding answers in the training data, which are written confidently. A hedge like *I am not sure, but* appears only where hedges appeared in similar text — not where the model's internal probabilities were low.

Internally, models do carry something like uncertainty: the probability assigned to each token. Research on this has a name, **calibration**, and models are moderately calibrated on multiple-choice questions and considerably worse on open text. But you cannot see those numbers in a chat window, and the model does not read them back to itself before writing. So confident phrasing carries almost no information about correctness.

> *Asking are you sure? is not a check. It is another prompt, and the most plausible continuation of a challenge is often an apology and a different answer, whether or not the first one was wrong.*

The two questions that do work

The useless move is to ask the model to grade its own answer. Two things do help.

Ask it to produce something checkable. Not *is this right* but *give me the steps*, *give me the DOI*, *write code that computes this*. A step, a DOI and a program can each be tested against something outside the conversation.

Ask it twice, cold. Open a new chat. Rephrase the question so it does not echo the first answer. Genuine general knowledge is stable across rephrasings; fabrication usually is not. If you get two different citations for the same claim, at least one is invented and probably both are.

What retrieval does and does not fix

Many tools now search the web and attach links. This helps, and it changes the failure rather than removing it. Now the model is summarising retrieved pages, so the summary can misstate a page that is itself correct, and the link can point at a real page that does not contain the claim. Studies of search-grounded assistants keep finding the same pattern: the citations exist, and a meaningful fraction do not support the sentence they are attached to.

The rule survives the upgrade. **Open the link.** If you did not open it, you did not check it.

Doing this without paying for anything

Everything above works on free tiers. For verification specifically, the free tools that matter are Google Scholar, the DOI resolver at `doi.org`, arXiv, PubMed, and Unpaywall's browser extension, which finds legal free copies of paywalled papers. Your library almost certainly has an institutional login too, and the librarian is the most underused expert on your campus.

The habit to build

When you read a model's answer, mentally underline every proper noun, number and reference. Those are the parts that need a source. The prose around them is usually fine. This takes about ten seconds and catches nearly everything that would otherwise cost you marks.

BEFORE YOU MOVE ON

A student gets a model's answer, replies "are you certain about that reference?", and the model confirms it. What has the confirmation established?

A. That the reference is probably real, since models are trained to admit uncertainty when they have it.

B. Almost nothing, because the follow-up is just another prompt and the model has no lookup to consult.

C. That the reference is real if the model gave the same answer both times, since consistency implies retrieval from memory.

D. That the model's confidence calibration was high for this answer, which is a reasonable proxy for accuracy on citations.

---

5 · 8 min

The fluency trap

THE ONE THING TO KEEP

Clarity produces the feeling of understanding before any of the work that earns it, so the only honest test of an AI explanation is whether you can reproduce it from a blank page afterwards.

Reading a clear explanation feels like understanding

Here is an experiment you can run on yourself in four minutes. Pick a topic you believe you understand — how a bicycle gear changes the effort you feel,

how a refrigerator makes cold, how a bill becomes law where you live. Rate your understanding out of seven. Now write the full causal explanation on paper, step by step, with no gaps. Then rate it again.

Almost everybody rates lower the second time. Psychologists call this the **illusion of explanatory depth**, from work by Rozenblit and Keil in 2002. We routinely mistake familiarity with a topic for the ability to explain it. The illusion is strongest for mechanical and procedural things, and weakest for facts and narratives.

AI assistants are illusion machines, not because they lie but because they are clear. A well-structured explanation with headings and a worked example produces exactly the sensation of having understood. The sensation arrives before any of the work that would earn it.

Fluency is a signal about the text, not about you

Cognitive psychology calls the underlying error **fluency misattribution**. When something is easy to process — clean typography, a familiar accent, a well-organised paragraph — we attribute the ease to our own grasp of the material. Studies going back to the 1990s show students rating their learning higher after reading a clear passage than after struggling with a harder one, then performing worse on the test.

This is why re-reading notes remains the most popular study method and one of the least effective. Re-reading raises fluency. Fluency raises confidence. Confidence does not raise the score.

A model raises fluency more than any textbook can, because it will rewrite the explanation until it feels effortless, adjusting to your exact confusion. That is genuinely valuable. It is also the moment to be most suspicious of your own sense of progress.

Desirable difficulties

Robert Bjork's phrase for the countermeasure is **desirable difficulties**: conditions that slow learning down while it happens and improve what remains afterwards. The best-established are:

- **Retrieval instead of review.** Recalling something from a blank page beats reading it again, by a wide margin, in dozens of experiments.
- **Spacing instead of massing.** The same total study time split across days beats one long block.
- **Interleaving instead of blocking.** Mixing problem types beats doing twenty of one type, even though it feels worse while you do it.
- **Generation before instruction.** Attempting an answer before being told, even wrongly, improves retention of the correct answer.

Every one of them feels worse than the alternative. That is the trap: the methods that feel productive and the methods that work are close to opposites. Students consistently rate massed re-reading as more effective than spaced retrieval and consistently score lower with it.

What this means when the model is helping

You can use these tools in a way that raises difficulty rather than lowering it. The whole of Module 3 is how. But the diagnostic is simple and you can apply it today:

> *After the model explains something, close the window and write the explanation from memory. If you cannot, you did not learn it. You read it.*

The gap between what you can read along with and what you can produce cold is the whole subject. It is also, precisely, the gap between the tutorial and the exam.

A cheap instrument for measuring the trap

Keep two columns in a notebook for a week. In the left, what you felt you understood after each AI-assisted session, out of five. In the right, what you could actually reconstruct the next morning before opening anything, out of five.

Most people find a consistent gap of one to two points. The size of your personal gap is more useful than any general advice, because it tells you how much to discount your own sense of having got it. Someone with a two-point

gap needs to build the checking habits urgently. Someone with a half-point gap already has them.

The one exception worth naming

Fluency is not always a lie. For material that genuinely is simple, or that you genuinely do already know, a clear explanation confirming it costs nothing and saves time. The trap is specific: it bites when the material is new, mechanical, and multi-step. Those are the topics where you must convert reading into producing before you believe your own confidence.

BEFORE YOU MOVE ON

Two students revise the same topic for an hour. One re-reads a model's clear explanation four times and feels confident. The other attempts the explanation from memory, fails, reads it once, and attempts again, ending the hour feeling shaky. What does the research predict for the exam a week later?

A. The shaky student scores higher, because retrieval attempts strengthen recall while re-reading mainly raises fluency.

B. Roughly equal scores, since both spent an hour and time on task is the dominant factor in retention.

C. The confident student scores higher, because confidence reduces exam anxiety and anxiety is the main cause of underperformance.

D. The shaky student scores higher only if their attempts were mostly correct, since retrieving errors reinforces the errors.

6 · 8 min

Context, limits, and why it forgets

THE ONE THING TO KEEP

Everything the model can use must fit in a fixed token window, it attends least to the middle of a long one, and quotas are measured in the same tokens — so pasting less and starting fresh chats improves answers and stretches a free tier at once.

The window

Everything a model can use to answer you must fit in its **context window** — a fixed number of tokens holding the system instructions, the whole conversation so far, any files you attached, and the answer being written. Nothing outside that window exists for the model.

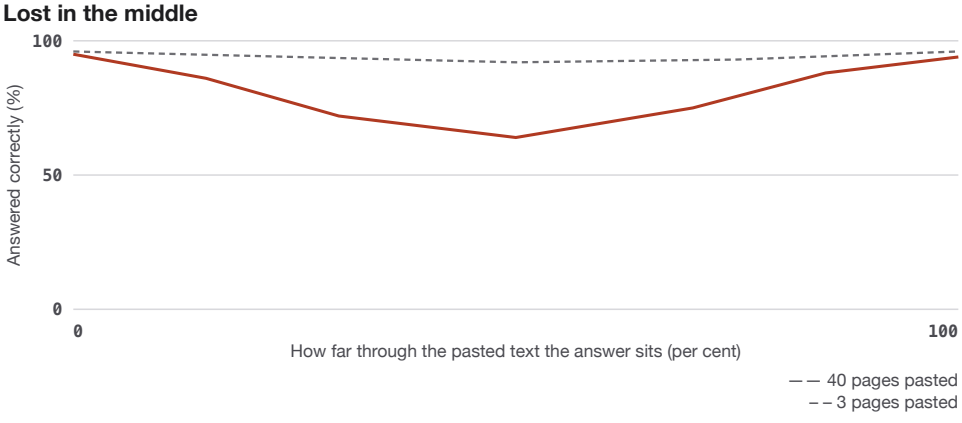
Sizes vary by model and change often. What matters is the arithmetic, not the current number. A 128,000-token window is roughly 90,000 English words, or a 300-page book. An 8,000-token window is roughly a long essay. Your uploaded 60-page PDF might be 30,000 tokens, and if the window is 8,000, most of it is gone before the model reads a word.

When the conversation exceeds the window, tools handle it in different ways, and none of them is *remembering*. Some truncate the oldest turns. Some summarise them into a shorter note and drop the originals. Either way, details from early in a long chat quietly stop existing, which is why an assistant that knew your essay's thesis in message four contradicts it in message forty.

Long context is not the same as attention

A large window means the text *fits*. It does not mean the model uses all of it evenly. Repeated experiments find a pattern researchers describe as *lost in the middle*: information at the very start and the very end of a long context is used reliably, and information buried in the middle is used much less. If you paste

a 40-page document and ask about something on page 22, expect worse answers than for page 1 or page 40.



One class exercise of sixty questions, in the shape repeated experiments keep finding. A large window means the text fits, not that it is used evenly: material at the very start and the very end is used reliably, and material buried in the middle much less. The fix is not a bigger window. It is pasting the three pages that contain the answer, which gives better replies, faster ones, and a free tier that lasts longer.

The practical fix is not a bigger window. It is putting less in. Paste the three relevant pages instead of the whole PDF. Quote the paragraph rather than the chapter. You will get better answers and faster ones.

Why free tiers throttle, in the same units

Providers charge and limit by tokens because tokens are what the computation costs. Every token in the window is processed for every token generated, so a long conversation costs more per reply than a short one — the cost of a chat grows faster than its length.

This is why:

- A free tier gives you a message allowance that empties faster when your messages are long.
- Uploading a large PDF can consume a surprising share of a daily quota in one action.
- The same tool feels more generous in English than in Hindi, Bengali or Tamil, because the same meaning costs more tokens in scripts the tokeniz-

er handles less efficiently.

None of this requires paying anything to work around. It requires being deliberate about what you put in the window.

Managing conversations on purpose

Three habits, all free:

1. **Start a new chat when the subject changes.** A fresh window with a clean statement of the problem beats a 60-message thread carrying three abandoned topics. Old context does not help; it distracts, because the model conditions on all of it.
2. **Re-state the constraints periodically.** In a long working session, every ten or so messages, restate the target: *reminder: 1,200 words, sceptical of the standard reading, A-level not undergraduate*. Cheap, and it survives truncation.
3. **Keep the source of truth outside the chat.** Your essay lives in a document. Your notes live in a file. The chat is a workspace you can throw away, not a place things are stored. If the conversation is your only copy of something, you will lose it.

Memory features

Some tools now keep a persistent memory across conversations — your name, your course, your preferences. This is convenient and has two consequences worth knowing. It is stored on the provider's servers, so it is subject to whatever their retention policy says. And it silently conditions future answers, so an assistant that *knows* you are a first-year can quietly simplify things you now need in full. Both are manageable; both are worth checking in the settings rather than assuming.

Running one on your own machine

If your laptop has 8GB of RAM you can run a small model locally and free with Ollama or llama.cpp — a 3B or 4B parameter model quantised to about 2-3GB. It is noticeably weaker than the hosted ones and it is entirely private,

works offline, and has no quota. For vocabulary drilling, rephrasing, or generating practice questions from text you supply, a small local model is genuinely enough. For anything needing broad knowledge, it is not. Knowing which of those you are doing is the skill.

BEFORE YOU MOVE ON

A student uploads a 200-page textbook and asks a question about a definition on page 96. The answer is confidently wrong, though the same model answers correctly when given just that page. Why?

A. PDF uploads strip formatting, so the model could not identify which text was a definition.

B. The model's training data did not include that textbook, so it fell back on general knowledge.

C. Material in the middle of a very long context is used far less reliably than material at the start or end.

D. The upload exceeded the daily token quota, so the tool silently answered from a truncated summary of the whole book.

---

7 · 9 min

Which tool for which job

THE ONE THING TO KEEP

Match the tool to the failure you can afford: deterministic tools for arithmetic, existence and data, and the model for explanation, phrasing, and deciding which deterministic tool to reach for.

The default is wrong more often than it is right

Opening a chatbot has become the reflex for every question, the way opening a search box was ten years ago. For a fair share of schoolwork it is the wrong in-

strument, and the wrongness is predictable from the mechanism in the first lesson.

Use this as a first pass.

Exact arithmetic or algebra. Not a chatbot on its own. A calculator, a spreadsheet, or a computer algebra system. Free and genuinely good: **SymPy** (Python, free), **Maxima** (free, decades old, powerful), **GeoGebra** (free, browser, excellent for graphing and geometry), **Desmos** (free graphing). Job ads and universities mention MATLAB and Mathematica; **GNU Octave** runs most MATLAB code free, and **SageMath** covers much of Mathematica's ground. If your chat tool has a code interpreter, that counts too, because it writes Python and actually runs it rather than predicting the answer.

Finding out whether something exists. Not a chatbot. A search engine or a database. Google Scholar, PubMed, arXiv, your library catalogue. A chatbot answers *what would a source about this look like*. A database answers *does this source exist*.

Current facts. Search, or a chatbot with search switched on and the links opened. Prices, timetables, exam dates, current office-holders, this year's syllabus.

Statistics on your own data. A spreadsheet (LibreOffice Calc, free; Google Sheets, free), or **jamovi** or **JASP** — both free, both give you SPSS-style output without SPSS's licence, both are what a lot of psychology departments now actually use. Ask the model which test and why; run the test in the tool.

Reading a scanned handout. OCR. Free options: Google Lens on a phone, **Tesseract** on a desktop. Many chat tools now read images directly, which is fine for short text and unreliable for tables and handwriting.

Writing and rewriting prose, explaining a concept, generating practice questions, translating, summarising text you supply, arguing with you. This is the chatbot's home ground. It is very good at all of it.

Storing anything. Not the chat. A document, a folder, a repository.

The free stack, named

Courses and job ads name paid tools because workplaces buy them. Learn the concept on the free one and the paid one takes an afternoon.

- Documents: Microsoft Word → **LibreOffice Writer**, Google Docs
- Spreadsheets: Excel → **LibreOffice Calc**, Google Sheets
- Statistics: SPSS, Stata → **jamovi**, **JASP**, **R** with RStudio
- Maths: Mathematica, MATLAB → **SageMath**, **GNU Octave**, **SymPy**
- References: EndNote → **Zotero** (free, and better)
- Diagrams: Visio → **draw.io**, **Inkscape**
- Images for coursework: Photoshop → **Photopea** (runs in a browser, free), **GIMP**, **Krita**
- Video for a presentation: Premiere → **DaVinci Resolve** (free tier is full-featured), **Shotcut**, **Kdenlive**
- Audio: → **Audacity**
- Notes: → **Obsidian**, **Joplin**, plain markdown files
- Programming: → Python, VS Code or VSCodium, Google Colab (free GPU hours)

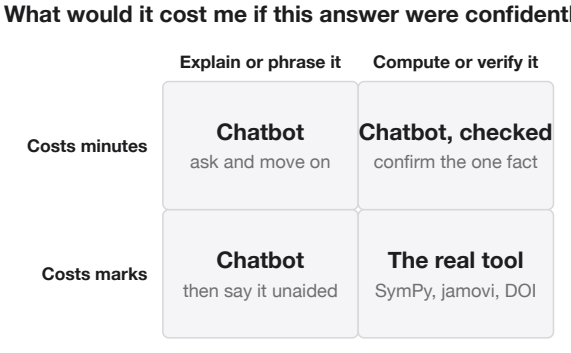
Every one of these runs on a modest laptop; several run on a phone browser.

Combining them is where the value is

The strong pattern is not one tool, it is a handoff. Ask the model *which statistical test suits this design and why*, then run it in jamovi. Ask it *give me the SymPy code to check this integral*, then run the code. Ask it *what would I search to find whether this claim has been tested*, then search Scholar. In each case the model does the part it is good at — knowing the shape of the field, phrasing the query, explaining the choice — and something deterministic does the part where being wrong is expensive.

The one-sentence test

Before opening a chat window, ask: **what would it cost me if this answer were confidently wrong?** If the answer is a wasted five minutes, use the chatbot and move on. If the answer 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 and let the model advise you about which one.



One question sorts most of schoolwork and needs no list to remember. The bottom right is where the free deterministic tools belong: SymPy or Maxima for the arithmetic, jamovi or JASP for the test, Scholar and the DOI resolver for whether a source exists. The model is still useful in that cell. It is good at saying which tool to open and why.

This single question sorts most of schoolwork correctly, and it does not require you to remember any list.

When you have only a phone

A great deal of this works on a phone browser: GeoGebra, Desmos, Photopea, Google Sheets, Colab, Zotero's web library, Scholar. What does not work well on a phone is anything requiring a large screen for comparison — reading a paper alongside your notes, or checking code output against a spec. If a phone is your only machine, use your institution's computer room for those tasks specifically, and do everything else where you are.

BEFORE YOU MOVE ON

A student needs to check whether the integral they computed by hand is correct before submitting. Which approach is structurally sound?

A. Ask the model to differentiate its own answer and confirm it matches the integrand, since that is a self-check.

B. Ask the model, then ask a second model, and accept the answer if both agree.

C. Ask the model to show every step of the integration, then read the steps carefully, since a visible chain of reasoning cannot hide an error.

D. Ask the model for SymPy or Maxima code, run it, and compare with the hand result.

8 · 8 min

Getting set up for nothing

THE ONE THING TO KEEP

Free tiers, Hugging Face Spaces, Colab, and a small local model on Ollama cover almost everything in this course, so the real skill is spending a limited allowance on the questions that are actually hard.

The constraint most advice ignores

A lot of study advice assumes a laptop with 16GB of RAM, a stable connection and a card on file. Plenty of students have a shared machine, an older phone, patchy data and no card. None of that stops you doing any of this course. It changes which route you take, and the routes are worth knowing before you need them.

Free access to capable models

Every major provider runs a free tier with a daily or hourly message cap. The caps move; the shape does not. Practical consequences:

- **Spend your allowance on the hard questions.** Do the easy rewriting with a weaker free model or on your own.
- **Batch.** One message containing five related questions costs one message from your allowance. Five messages cost five.
- **Keep two accounts on different providers.** When one runs out, switch. This is not a trick, it is what the free tiers are for.

Beyond the big names: **Hugging Face Spaces** hosts thousands of free demos, including chat interfaces to open-weight models, and needs no payment. **Google Colab** gives free GPU time in a notebook, enough to run a mid-sized open model for an afternoon. Some universities provide staff and students with a paid tier free — check before assuming; a surprising number do and never announce it well.

Running a model on your own machine

This is the option that removes quotas, works offline, and keeps everything private. It needs RAM more than anything else.

- **8GB RAM:** a 3B or 4B parameter model, quantised to 4 bits, around 2-3GB on disk. Usable for rephrasing, practice questions from supplied text, vocabulary, simple code.
- **16GB RAM:** a 7B or 8B model, 4-5GB. Noticeably better; still well short of the hosted ones on hard reasoning.
- **Apple silicon** punches above its weight here because the memory is shared with the GPU.

The two free tools are **Ollama**, which is a single install and then `ollama run <model>` in a terminal, and **llama.cpp**, which is more work and runs on almost anything. **LM Studio** gives a graphical interface if terminals are unwelcome.

Be honest about the gap. A 4B model run locally will hallucinate more, follow multi-step instructions worse, and give up on long documents. For a set of tasks — drilling, rewriting, quizzing you from a chapter you paste in — it is entirely adequate, and the fact that it works on a train with no signal is worth a lot.

A phone-only setup that actually works

1. Chat tools: all major ones have web interfaces that work in a mobile browser; no app needed.
2. Documents: Google Docs or the free Microsoft web apps; both save continuously, which doubles as your process record.
3. Notes: Obsidian and Joplin both have mobile apps and sync free via a folder.
4. Maths: GeoGebra and Desmos are excellent on a phone.
5. Reading papers: get the PDF into a reader that lets you highlight; Zotero's mobile app is free.
6. Voice: dictation is free on every phone, and talking through a problem out loud, then pasting the transcript, is a genuinely good way to think.

The thing to avoid on a phone is writing long prose directly into a chat window. It encourages accepting whatever comes back because editing is painful. Draft in a document, paste in, paste back.

Data, and working offline

If data is expensive or intermittent:

- Download the material once. Lecture slides, past papers, PDFs. Read offline.
- Batch your questions. Write them in a note during the day, ask them together when you have signal.
- A local model needs no connection at all after the download.
- Wikipedia has offline packages via **Kiwix** — the whole of English Wikipedia without images is a few tens of gigabytes, and a subject-specific selection is far smaller.

What not to spend money on

There is a large market in AI study apps charging monthly for a wrapper around a model you can use free. Before paying for anything, ask what it does that a free chat window plus a document does not. Sometimes the answer is

real — a genuinely good spaced-repetition scheduler, for instance, though **Anki** is free on desktop and Android and is still the best one. Usually the answer is a nicer interface around the same free capability.

The honest exception is a paid tier of a major model when you are doing heavy work in a specific week. It is bought by the month and cancellable, and one month during a dissertation is a different decision from a standing subscription all year.

BEFORE YOU MOVE ON

A student with an 8GB laptop and no budget installs a 4-bit 3B model with Ollama. Which task is it genuinely well suited to?

A. Checking whether a set of references from an essay actually exist.

B. Generating practice questions from a chapter the student pastes in, offline.

C. Answering questions about last month's changes to the exam syllabus.

D. Explaining an unfamiliar advanced topic that is not in the student's materials, since local models are trained on the same data.

---

9 · 7 min

A working agreement with yourself

THE ONE THING TO KEEP

Write down in advance the five or six capabilities you will never outsource, your disclosure sentence, your verification rule and your evidence habit, because the decision made calmly is the one that holds at midnight.

Decide before you are tired

Every hard call about using these tools arrives at the worst possible moment: eleven at night, deadline tomorrow, three hundred words short. Nobody

makes good decisions there. So the decision has to be made now, in writing, while nothing is at stake.

This lesson asks you to write four short things. It takes fifteen minutes and it is the most useful fifteen minutes in this module.

One: the list of things you will not outsource

Name the specific capabilities your degree or your career is supposed to leave you with. Not subjects. Capabilities.

A medical student might write: *taking a history, reading an ECG, calculating a paediatric dose*. A computer science student: *writing a function from a spec without help, reading a stack trace, reasoning about complexity*. A history student: *reading a primary source and forming an argument about it, structuring a paragraph*. A language student: *producing a sentence in the target language without a prompt*.

These are the things you do unaided, always, even when it is slower. The list should be short — five or six items — because a long list is one you will abandon. Everything not on the list is fair game.

The test for whether something belongs on the list: **if you cannot do it, can you tell when somebody else has done it badly?** Supervision is most of what is left of many jobs, and you cannot supervise a skill you never acquired.

Two: your disclosure default

Decide now what you will write in a disclosure statement, and decide it on the generous side. Something like:

> *AI use: I used a language model to generate practice questions, to critique my draft structure, and to check my understanding of X. All prose, all argument and all sources are my own; I opened every source cited.*

Writing the sentence in advance changes behaviour, because a use you would be embarrassed to disclose is a use you now notice yourself making. Module 7

covers institutional rules in detail. This is the personal version, which binds you before the rules do.

Three: your check for specifics

One line: *every name, number, date, quotation and reference gets verified against a source I opened.* Written down, it is a rule. Unwritten, it is an intention, and intentions lose to deadlines.

Four: your evidence habit

Decide how you will be able to show your work is yours, before anyone asks. The cheapest version: write in a tool that keeps version history — Google Docs, Word with autosave, or plain files in git — and never paste a finished draft into a blank document. That is the whole habit. Module 7 explains why it matters more than any argument you could make after an accusation.

Why an agreement beats a rule

Institutional rules tell you what is punishable. They are necessarily crude, they differ between departments, and they are silent on most of what you do. An agreement with yourself covers the enormous space the rules do not reach: the moment you nearly asked for the whole paragraph, the citation you nearly did not open, the problem set you nearly copied because it was late.

Nobody will check this document. That is the point. The students who come out of the next few years able to do things are the ones who decided, in advance and for their own reasons, what they were going to keep doing themselves.

A worked example, because abstractions are easy to nod at

Here is a real-shaped agreement, written by a second-year engineering student:

> *Never outsourced: free-body diagrams, unit and dimension checks, writing a function from a spec, reading a datasheet, the argument of any*

> *report.*

> *Disclosure default: I will state which sections had AI critique and which had AI drafting, and there will be no sections with AI drafting.*

> *Specifics rule: every constant, standard number and reference gets opened. If I cannot open it, it comes out of the report.*

> *Evidence: everything in git, committed at the end of every session, even when the session was bad.*

Notice what it does not say. It does not ban the tool. It does not promise heroic restraint. It names four small mechanical commitments that can be checked by looking, which is the only kind of commitment that survives a bad week.

Revisit it

Put a date on it and read it again at the end of each term. The list should change. Things you deliberately learned unaided move off it, because you have them; new things move on. A list that has not changed in a year is a list you stopped reading.

BEFORE YOU MOVE ON

What is the stated test for whether a capability belongs on your do-not-outsource list?

A. Whether it appears in the assessment criteria for your course this year.

B. Whether an AI tool currently does it badly, since those are the skills that remain valuable.

C. Whether losing it would leave you unable to judge somebody else's work on it.

D. Whether doing it manually is slower than asking, since the time saving measures how much the tool is really contributing.

CASE STUDY · MODULE 1

The tutorial sheet that everyone finished

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.

STOP HERE

Before turning the page: what would you have done, and what would it have cost if you were wrong? Write it down. The point of the next section is to compare.

WHAT HAPPENED

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 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.

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.

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: answers and the reason behind each are at the back. If two questions from the same module go wrong, re-read that module before the exam rather than after.

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

MODULE 2

Asking well

The difference between a useless answer and a useful one is usually the question. Task, level, constraint, form; the source supplied rather than remembered; the follow-up that finds the weak joint.

> BY THE END OF THIS MODULE YOU CAN
>
> Write a study prompt that states the task, the level, the constraints and the required form, ground it in a source you supply rather than the model's memory, and diagnose from a bad answer whether the fault was the model's or your instruction's

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:

Four things the model cannot guess

Explain enzyme inhibition.	The same question with the four supplied
• No level, so it starts at the beginning	• Level: second year, reads a Lineweaver-Burk plot
• No task, so it tours the topic instead of answering a question	• Task: what happens to Vmax and Km, and why
• No constraints, so it runs to a page and defines what you know	• Constraints: 250 words, do not define substrate
• No form, so it arrives as a balanced mid-length essay	• Form: prose, ending in one question that would catch me
• You get the textbook page you have already read	• 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.
4. **The form** — prose, list, table, worked example, question.

Say what you already know, especially the wrong bits

The highest-value sentence in most study prompts is a statement of your current, possibly mistaken, understanding.

> *My current understanding is that a competitive inhibitor lowers Vmax because it blocks the active site. Tell me where that is wrong and why, and do not just give me the correct version.*

This works because it converts a general explanation task into a diagnosis task. The model now has something specific to act on, and you get the correction aimed exactly at your error rather than a tour of the topic that happens to contain the answer somewhere.

It also gives you a much better test of whether you have learned anything, because you wrote down a claim before reading the answer. Generating an answer before being told is one of the reliable memory effects in the literature; writing your wrong version into the prompt gets it for free.

Ask for the shape you actually need

Schoolwork has shapes. Name them.

- *Give me a worked example with the reasoning shown at each step, then a second, similar problem with the answer hidden.*
- *Give me the three strongest objections to this argument, strongest first, in one sentence each.*
- *Give me a five-line summary and then the part I would be most likely to misunderstand.*
- *Rewrite this paragraph in my own register without adding any new claims, and list what you changed.*

Each of these is a different tool. A prompt that does not name a shape gets the default shape, which is a mid-length balanced essay, which is almost never what a study session needs.

The negative constraints do most of the work

Models over-explain, hedge, and pad with caveats. This is a consequence of how they were tuned, not a fault you can argue them out of, but you can instruct around it.

Useful lines to keep:

- *No introduction, no summary, start with the substance.*
- *Do not tell me it depends. Give the standard case, then name the exception in one line.*
- *Assume I have read the chapter.*
- *Do not apologise or restate my question.*
- *If you are not confident about a specific figure, say so instead of giving one.*

That last one does not make the model reliable — it has no honest confidence signal in the sense of the previous module — but it does shift the style towards hedging on specifics, which makes the risky parts easier to spot.

Iterate rather than restarting

When an answer misses, the fix is usually one sentence, not a rewritten prompt. *Too general — do it for the specific case of a non-competitive inhibitor at saturating substrate. Too long — cut it to the third of that which I could not have got from a textbook. You defined terms I said I knew — try again without those.*

The habit worth building is noticing **which of the four things was missing**. Answer too basic: you omitted level. Answer wandered: you omitted the task. Answer too long: you omitted constraints. Answer in the wrong format: you omitted the form. Almost every disappointing answer maps onto one of those four, and diagnosing it takes five seconds.

The prompt is not the skill

Providers have spent enormous effort making these systems tolerant of vague prompts, and they are much better at it than they were. So do not turn prompt writing into a craft with rituals. Four items and a willingness to say *no, like this* covers essentially everything. The skill worth having is knowing what you actually want, which is a thinking problem, not a typing one.

BEFORE YOU MOVE ON

A student asks for an explanation of recursion and gets a long answer that begins by defining a function and a variable. Which element of the prompt was missing?

A. The level and prior knowledge, so the model started from the beginning rather than from where the student was.

B. The form, because they did not ask for a worked example rather than prose.

C. The task, because recursion is a topic rather than a job to be done.

D. The constraints, because no length was given and the answer therefore expanded to fill the default response size.

---

11 · 9 min

Give it the source

THE ONE THING TO KEEP

Paste the smallest span of the real source rather than relying on the model's memory, then check the summary against the passage instead of checking whether the fact exists.

Supplied text beats remembered text

The previous module's mechanism has a direct practical consequence. Facts the model recalls come from a lossy compression of enormous amounts of

text. Facts you paste into the window are just there, exactly as written. Anything you can supply, you should supply.

This one change fixes most of what students complain about. *It got the definition slightly wrong*. Paste the definition. *It used a different notation from my lecturer*. Paste the lecture slide. *It answered about the general case and my course does it differently*. Paste your course's version.

The pattern

> *Here is the relevant section of my lecture notes, between the markers.*
>
> *---*
>
> *[paste]*
>
> *---*
>
> *Using only what is in that text, explain X. If the text does not contain enough to answer, say which part is missing rather than filling it in from elsewhere.*

Three things are doing work here. The markers stop the model confusing your source with your instruction. *Using only what is in that text* narrows it to the supplied material. The final clause gives it a permitted way to fail, which makes silent invention less likely — though not impossible, so you still check.

Grounding does not equal correctness

Be precise about what this buys. Grounding on your source means the model is now summarising, restructuring and explaining a specific text rather than reconstructing one. It removes the *invention* failure and leaves two others:

- **Misreading.** It can still misstate what the passage says, particularly for negations, conditions and exceptions. *Unless the sample is randomised* is exactly the kind of clause that goes missing in a summary.
- **Wrong source.** If your lecture notes contain an error, or you pasted last year's version, the model will faithfully explain the error.

So the check after grounding is different from the check without it. You are no longer asking *does this exist*. You are asking *does the passage actually say that*, which you can settle in ten seconds by looking at the passage.

What to paste, and how much

Less than you think. The *lost in the middle* effect from Module 1 means a 40-page upload gets worse use of page 22 than a 2-page paste of the relevant section. Practical rule: paste the smallest span that contains the answer plus enough around it to make sense. A definition plus the paragraph it sits in. A method plus its assumptions. A code function plus the type it takes.

When you genuinely need a whole document — a paper you are reviewing, a chapter you are revising — upload it, and then work section by section rather than asking global questions. *Summarise section 3* is far more reliable than *summarise the paper*, and you can check it.

Where the text comes from when you have none

Students often say they have no source to paste. Usually they do:

- Lecture slides, exported as text or screenshotted (most tools read images now).
- The module handbook and the assessment criteria — these are extremely useful to paste and almost nobody does it.
- The textbook, photographed a page at a time. Free OCR if the image reading is poor: Google Lens on a phone, Tesseract on a desktop.
- Past papers and mark schemes.
- Your own earlier work, when you want continuity of terminology.

A related move that costs nothing: paste **the mark scheme** alongside your draft and ask which criteria the draft currently fails. That is grounding applied to assessment, and it is one of the highest-yield uses in this whole course.

Copyright and the practical position

Pasting a page of a textbook into a chat window to help you study is a use most people would consider ordinary, and in many countries some form of research or private-study exception exists. But copyright law differs sharply between jurisdictions, the exceptions have real limits, and whether feeding copyrighted text into a commercial AI service falls inside them is genuinely unsettled and being litigated in several countries at once. This course is not going to pretend it is resolved, and none of this is legal advice.

What is clearly sensible: keep it to the extent you need, do not upload entire books, do not redistribute what comes back as though it were yours, and check whether your institution's licence for its digital materials says anything about third-party services. Many do, and it is usually one paragraph.

The privacy half

Everything you paste goes to somebody's servers. That is fine for a lecture slide and not fine for a classmate's draft, a patient case from placement, an unpublished dataset, or anything with somebody's name and details in it. The next lessons in this module deal with that properly. For now, hold the rule: **supply generously from published or your own material, and never supply other people's private information.**

BEFORE YOU MOVE ON

A student pastes a lecture handout and asks the model to explain a method from it. The explanation is fluent but omits the condition that the method only applies to normally distributed data. What does this failure show?

A. That grounding failed, and the model answered from training memory instead of the supplied text.
B. That the handout should have been uploaded as a file rather than pasted, so the model could index it properly.
C. That the context window was exceeded, pushing the qualifying sentence out of what the model could see.
D. That grounding removes invention but not misreading, and conditions and exceptions are what summaries drop.

12 · 8 min

Ask for the working, then check it

THE ONE THING TO KEEP

Ask for steps because they improve accuracy and give you something auditable, then audit with units, magnitude and the one difficult joint — never treating the stated reasoning as proof that it is what produced the answer.

A conclusion cannot be checked; a chain can

If a model tells you the answer is 4.2 kJ, you have one thing to verify and no way to verify it short of doing the problem yourself. If it gives you six steps, you have six small things to verify, most of which take seconds, and an error anywhere is usually visible.

This is the practical reason to ask for working. There is a second reason: models are measurably more accurate when they produce intermediate steps. Generating reasoning before the answer improves performance on multi-step

problems, and the effect is large enough that it is now built into how most assistants respond by default. Asking for the steps improves the answer and gives you something to inspect. It costs one clause.

> *Work through it step by step. State each assumption as you use it. Give the final answer last, on its own line.*

Steps are not proof

Here is the honest limitation, and it matters. A chain of reasoning produced by a model is *also* generated text. It reads as the derivation that produced the answer; it is a plausible derivation written alongside it. Research on this — often discussed as the faithfulness of chain-of-thought — repeatedly finds cases where the stated reasoning does not match what actually drove the answer, including cases where a model given a hint changes its answer to match the hint while producing a chain that never mentions it.

So visible working buys you two real things:

1. **Checkability.** You can find the wrong step yourself.
2. **Accuracy.** Step-by-step generation genuinely does better on multi-step problems.

It does not buy you a guarantee that the reasoning shown is the reasoning used. Do not treat a plausible chain as verification. Treat it as a document to audit.

How to audit a chain quickly

You rarely need to check every step. Three passes, in order, catch nearly everything:

Units and dimensions. In any physical or engineering problem, carry the units through. If joules appear where you expected joules per second, the error is upstream of that line and you can stop reading. This one check catches an enormous proportion of numerical errors and takes about fifteen seconds.

Magnitude. Is the answer the right size? A drug dose of 4 grams where you expected milligrams, a distance of 300 metres where you expected kilometres, a probability of 1.4. Estimate roughly before you look at the answer so you have something to compare against.

The joint. Most wrong chains are right for several steps and then make one unjustified move — a cancellation that requires a term to be non-zero, an approximation applied outside its range, a rule from one case used in another. Find the step where the difficulty of the problem actually lives and read that one properly.

Auditing a chain, about two minutes

A conclusion gives you one thing to verify and no way to verify it. Six steps give you six small ones, most of which take seconds. None of this proves that the reasoning shown is the reasoning that produced the answer, and research on the faithfulness of these chains keeps finding cases where it is not. Visible working is a document to audit, not a guarantee.

Make it show the assumptions separately

A useful variation:

> *Before solving, list every assumption the solution will rely on. Then solve. Then say which assumption, if wrong, would change the answer most.*

This surfaces the thing you actually need to check. In coursework, the assumption list is often worth more than the answer — it is frequently exactly what the marker is looking for, and it is the part students omit.

The re-derivation test

The strongest cheap check: cover the chain and redo the step you are least sure about yourself. If you can reproduce it, it is right and you have practised. If you cannot, either it is wrong or you have found the thing you need to learn. Both outcomes are useful, which is unusual for a check.

This is also the point where a lot of the learning happens. Reading somebody else's correct derivation teaches you very little. Regenerating one step of it teaches you a lot, for the reasons in the fluency lesson.

For non-numerical subjects

The same move works outside the sciences, in a different vocabulary.

- *Set out the argument as numbered premises and a conclusion before writing the paragraph.* Then check whether the conclusion actually follows, and whether any premise is doing work it cannot support.
- *Which of these claims is contested among historians, and by whom?* Then check that the contest is real.
- *Which sentence in this paragraph carries the most weight, and what would have to be true for it to hold?*

In every case you are converting a smooth output into a set of joints you can test individually. That is the whole technique.

BEFORE YOU MOVE ON

Why is a step-by-step chain from a model not evidence that the reasoning shown is the reasoning that produced the answer?

A. Because the temperature setting randomises intermediate steps more than final answers.

B. Because models generate the final answer first and then write a justification backwards to fit it.

C. Because the chain is generated text written alongside the answer, not a record of what produced it.

D. Because chains are usually copied from training examples of similar problems rather than computed for this one.

13 · 8 min

Follow-ups that find the weak joint

THE ONE THING TO KEEP

Replace "are you sure" with follow-ups that produce checkable material — what would make this wrong, argue the opposite, what would a marker criticise — and test stability by re-asking cold in a fresh chat.

Stop asking whether it is sure

Are you sure? is the most common follow-up and one of the least useful. It is another prompt, and the model has no lookup to consult. Worse, the tuning that makes assistants agreeable means a challenge often produces a retraction whether or not the original was wrong. You can talk a correct model out of a correct answer, which should tell you exactly how much information the retraction carries.

So replace it. The follow-ups below are the ones that actually move you forward, because each asks for something that can be checked outside the conversation or that changes the shape of the answer rather than its confidence.

The five that earn their place

1. What would make this wrong?

> *What would have to be true for this answer to be incorrect? What is the strongest case against it?*

This converts an assertion into a claim with stated conditions. It surfaces the assumptions and it works in every subject. In an essay it gives you your counterargument paragraph; in a lab report it gives you your limitations section.

2. Which part of this are you least confident about?

Not a real confidence report, for reasons already covered. But it reliably points at the parts that are unusual, contested or specific, because those are what surrounding text hedges about. Treat the answer as a list of things to verify, not as a probability.

3. Argue the opposite, as strongly as you can.

The best single move for essay subjects. Ask for the strongest version of the position you disagree with, written by somebody who believes it. Then look at whether your own argument survives it. If the opposing case is easy to write and hard to answer, you have picked the wrong thesis, and it is much better to find that out now.

4. What would a marker in this subject criticise?

Better with the mark scheme pasted in, but useful even without. Framing the model as an examiner rather than a helper changes the output substantially, because it shifts from completion to critique.

5. What am I not asking that I should be?

Works surprisingly well when you are new to a topic and do not know its shape. It surfaces the standard next questions in a field. Verify anything specific that comes back, as always.

The rephrase-and-repeat check

When a specific answer matters, do this:

1. Open a new chat.
2. Ask the same question in different words, without pasting the earlier answer.
3. Compare.

Stable answers across cold rephrasings are more likely to reflect well-represented knowledge. Divergent answers mean you need a real source. It costs one message and it is the closest thing to a free reliability test you have.

The reason it must be a new chat: everything in the current context conditions the next token strongly. A model that has just written an answer will tend to be consistent with it. Consistency inside a conversation measures almost nothing.

Asking it to disagree with itself

A structured variation worth knowing:

> *Here is an answer to a problem. Act as a sceptical examiner. Find the weakest step. Do not be polite about it.*

Paste the previous answer into a *fresh* conversation rather than continuing the old one. Stripped of the context in which it produced the answer, the model critiques much more freely. This is not magic and it will sometimes invent objections, but the objections are cheap to evaluate and it catches real errors often enough to be worth the message.

The follow-up that is actually about you

The last one is not addressed to the model:

> *Can I now explain this without the window open?*

If yes, the session worked. If no, you have a fluent answer and no new capability, and the right move is to close it and try to reconstruct from memory before reading again. Everything in Module 3 is built on that question.

What to do when it will not budge

Sometimes a model repeats the same wrong thing however you ask. This usually means the error is well-represented in its training data — a common misconception, an outdated standard, a widely-copied wrong answer. That is a signal, not a nuisance. Go and find out why the wrong version is so common; it is often the exact thing your marker is testing whether you know.

BEFORE YOU MOVE ON

Why must the rephrase-and-repeat reliability check be done in a new conversation rather than further down the same one?

A. Because a long conversation eventually exceeds the context window and the original question is truncated away.
B. Because the earlier answer sits in the context and heavily conditions what follows, so consistency proves little.
C. Because models apply a lower temperature to later messages in a conversation, reducing variation artificially.
D. Because a repeated question in the same chat is treated as a challenge, and the tuning makes the model retract rather than repeat.

14 · 8 min

Explain it at three levels

THE ONE THING TO KEEP

Ask for the same idea below your level, at your level, and above it — the third pass names which parts of what you were taught are teaching approximations, and that is usually what separates a good answer from an excellent one.

Difficulty is a dial you can set

One of the genuinely new things these tools offer is control over the level of an explanation. A textbook is written at one level for everybody. A model will write the same content at any level you name, as many times as you like, and switching between levels is where a lot of the understanding actually happens.

The technique: ask for the same idea three times, deliberately.

1. **At the level below yours.** Simple language, no jargon, an everyday analogy. This is where you find out whether you can hold the shape of the

idea at all.

2. **At your level.** Correct terminology, the standard treatment, the notation your course uses.
3. **At the level above.** What the simplification hid, where the standard treatment breaks, what a specialist would object to in the version at level 2.

The third one is the valuable one and almost nobody asks for it. It tells you which parts of what you learned are true and which are teaching approximations, and that distinction is frequently exactly what separates a good answer from an excellent one in an exam.

An example

For *why does ice float*:

- Level below: water molecules line up in a lattice with gaps when they freeze, so the same amount of water takes more space.
- Your level: hydrogen bonding forces a tetrahedral arrangement in ice Ih, lowering density relative to liquid water at 0 degrees; density maximum at about 4 degrees.
- Level above: this is specific to ice Ih at ordinary pressure; there are more than a dozen other ice phases, several denser than liquid water, and the density anomaly is a consequence of the directional bonding, not a general property of solids.

The third level is where you learn that *solids are denser than liquids, except water* is itself a simplification.

Analogies, and their expiry date

Ask for an analogy and you will get one. They are useful and they all break. The important habit is to ask where.

> *Give me an analogy for how a neural network learns. Then tell me exactly where the analogy stops being true and what someone would get wrong if they took it too far.*

Every analogy in science teaching has a failure point that a marker will probe. Electrons orbiting like planets: fails at quantisation and at what an orbital actually is. The cell as a factory: fails on scale, on the absence of a manager, and on the fact that everything is bumping around at random. Getting the failure point in the same breath as the analogy stops it becoming a misconception you then have to unlearn.

The Feynman move, and its honest version

The well-known technique is: explain the topic in plain language as if to somebody who does not know it, notice where you stumble, go back to the source for those bits, repeat.

The AI version people usually describe is *ask the model to explain it simply*. That is the wrong half. The valuable half is **you** doing the explaining. The correct use is:

> *I am going to explain X to you as if you are a bright 15-year-old who knows no biology. Do not help me. Afterwards, tell me: which parts were vague, which were actually wrong, and which I skipped over because I do not understand them.*

Now the model is a listener and a critic, and you have done the production. This is one of the best single uses of a language model in study, and it inverts what most students do with it.

Levels for reading, too

The same dial works on material you are struggling with. Paste the difficult paragraph and ask for it at the level below, then read the original again. The point is not to replace the original — it is to get enough of a foothold that the original becomes readable. Then read the original.

This matters for dense primary material: a paper's methods section, a legal judgment, a philosophical text. The simplified version is a ladder, not a destination, and the thing you will be examined on is the original.

The limit

Level control is only as good as the model's grasp of the subject, and at the level above it is generating what a specialist objection sounds like. On mainstream material that is reliable enough to be useful. On the frontier of a narrow field it is not, and a research supervisor will spot the difference immediately. Use level three to know what questions to ask, then ask a human.

BEFORE YOU MOVE ON

What makes the third pass — the explanation above your level — the most valuable of the three?

A. It gives you vocabulary that markers reward, which raises the register of your written answers.

B. It covers material that will appear in later years, so you are working ahead of the syllabus.

C. It is the level at which the model is most accurate, because advanced text is better represented in scientific training data.

D. It names which parts of the standard treatment are teaching approximations rather than facts.

15 · 8 min

Conversations have a shape

THE ONE THING TO KEEP

A conversation is a growing block of text the model re-reads every turn, so drift, anchoring and agreeableness accumulate — carry a short summary into a fresh chat instead of arguing with a long one.

Long threads go bad in predictable ways

A conversation is not a relationship with an assistant that gets to know you. It is a growing block of text that the model re-reads in full before every reply. That single fact explains everything that goes wrong in long sessions.

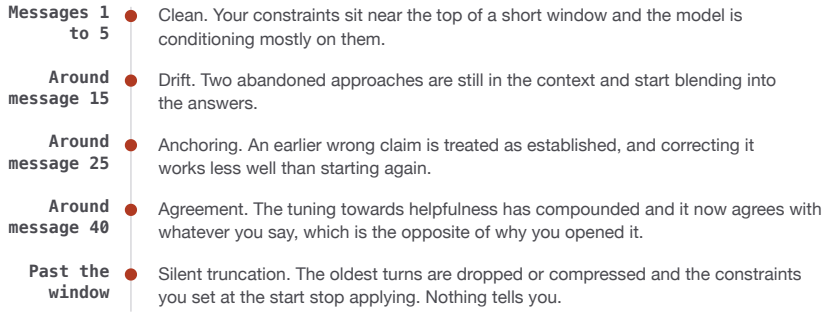
Drift. Forty messages in, the model is conditioning on forty messages of context, including three abandoned approaches and a tangent about something else. Its answers start blending them. You get an essay plan that quietly reverts to a structure you rejected twenty minutes ago.

Anchoring. Once a wrong claim is in the context, it tends to persist. The model treats its own earlier output as established. Corrections help less than starting again, because the wrong version is still sitting there being conditioned on.

Sycophancy compounding. Assistants are tuned to be agreeable. Over a long thread of you pushing back, that tuning accumulates, and you end up with a model that agrees with everything you say. At that point it has stopped being useful as a critic, which is usually why you opened it.

Silent truncation. When the thread exceeds the window, the oldest content is dropped or compressed. Nothing tells you. Constraints you set at the start stop applying.

How a long thread goes bad



None of this is the model getting tired. It re-reads the whole conversation before every reply, so everything in the thread conditions the next answer, including the parts you abandoned. The repair costs fifteen seconds: ask for the conclusions and constraints in under 150 words, check that summary, and paste it into a fresh chat.

The fix is cheap and unglamorous

Start a new chat far more often than feels natural. New subject, new chat. Wrong turn, new chat. Twenty messages in, new chat. Copy forward only the two or three things that matter: the current draft, the constraint list, the conclusion you reached. This takes fifteen seconds and it is the single highest-value habit in this module.

A useful ritual before closing a long thread:

> *Summarise, in under 150 words, the conclusions we reached and the constraints I set. I am going to paste this into a fresh conversation.*

Then check the summary is right, and paste it into a new window. You have kept the signal and dropped the accumulated noise.

Restate constraints periodically

In a working session, every ten messages or so:

> *Reminder: 1,500 words, undergraduate history, argument is that the reforms were driven by fiscal pressure rather than ideology, British spelling, no new sources.*

Cheap insurance against truncation and drift, and it forces you to keep the constraints explicit in your own head.

One thread, one job

Keep separate conversations for separate jobs, in the way you would keep separate documents:

- One for the essay you are writing.
- One for the problem set.
- One for the reading you are trying to understand.
- One scratch thread you throw away daily.

Mixing them costs you quality because everything conditions on everything else. Separating them also makes your process record legible later, which matters when you write a disclosure statement.

Custom instructions and saved prompts

Most tools let you set standing instructions that apply to every new conversation. This is where the boilerplate belongs, so you stop retyping it:

> *I am a second-year undergraduate in economics. British spelling. Be direct; no preamble, no summaries of my question, no offers of further help. When I ask about a specific figure, source or paper, say plainly if you are not confident rather than producing one. When I ask you to check my reasoning, look for the weakest step rather than confirming.*

Two warnings. First, standing instructions are also standing context — a note that you are a second-year will simplify things you later need in full, so revisit it. Second, whatever you write there is stored on the provider's servers along with any memory feature, so keep it professional and impersonal.

Beyond that, keep a plain text file of prompts that worked: the quiz-me prompt, the marker prompt, the explain-back prompt. A file of six good prompts you actually reuse beats any purchased prompt library.

The signal that a thread is finished

When you find yourself explaining to the model something it already agreed to twice, the thread is over. Do not argue with it. It is not being stubborn; it is being conditioned on a context that now contains the argument. Close it, start again, and give the fresh window a clean statement of where you got to.

BEFORE YOU MOVE ON

Thirty messages into an essay session, the model keeps reintroducing a structure the student rejected early on. What is the most effective response?

A. Ask it to summarise the conclusions and constraints, then paste that summary into a new conversation.

B. Tell it firmly to stop using that structure and to remember the correction this time.

C. Switch to a model with a larger context window so the early rejection is no longer truncated.

D. Delete the messages containing the rejected structure using the tool's edit feature, so the context no longer includes it.

---

16 · 9 min

What not to paste

THE ONE THING TO KEEP

You can consent to a service handling your own data but never somebody else's, so placement, participant and classmate material stays out of the window regardless of privacy settings.

The window is not private

What you type goes to a company's servers. It is stored. Depending on the product and your settings, it may be used to improve the service, which can

mean human review of samples. Consumer free tiers are the most likely to train on your input; paid business and education accounts usually contract not to. The settings differ by provider and change, so the only reliable move is to look at yours once and know what it says.

That is the background. The foreground is a short list of things that should not go in, whatever the settings say.

The list

Other people's personal information. A classmate's draft, a group member's contact details, a screenshot of a chat, an interview transcript with names in it. You may consent to a service handling your data. You cannot consent on somebody else's behalf.

Anything from a clinical, legal or social-care placement. Patient details, case notes, safeguarding information — including de-identified accounts that a person could still be recognised from. Medical and nursing students hit this constantly and the answer is always the same: it does not go in. Discuss the *clinical question* in general terms if you need to. Not the patient.

Research data with human participants. Your ethics approval said where the data would be stored and who would see it. A commercial AI service is almost certainly not on that list. Uploading it can be a genuine research-integrity breach, independent of any AI policy.

Unpublished work that is not yours. A supervisor's draft paper, a collaborator's data, a company's material from a placement. Placements often come with a confidentiality agreement that you signed and have half forgotten.

Credentials and identifiers. API keys, passwords, student ID numbers, bank details, passport scans. Obvious, and it happens weekly, usually inside a pasted config file or a screenshot with a browser tab visible.

Your own most sensitive material. Anything about your health, your finances, your immigration status, an appeal you are making, a disciplinary matter. There is no rule against it and you may decide it is worth it. Decide deliberately, knowing it is stored, rather than by default.

Redaction that actually works

When you need help with something that contains details you should not share, strip it first.

- Replace names with roles: *Patient A*, *the supervisor*, *Company X*.
- Change or remove dates, locations and ages where they are not load-bearing. If the age matters clinically, band it: *late 60s*.
- Remove anything that is unique in combination. A rare condition plus a small town plus a date identifies somebody even with the name gone. This is the failure most people miss, and it is why simple name removal is not de-identification.
- For code: replace keys with `REDACTED`, and if a key has ever been pasted anywhere, rotate it. Assume anything pasted is compromised.

Institutional accounts

If your university provides an AI tool, use it for coursework. Two reasons. The contract usually forbids training on your input, and the data usually stays inside an agreement your institution has already assessed. It is also, often, a paid tier for free.

The trade-off is worth knowing: an institutional account may be visible to your institution in ways a personal one is not. Administrators generally cannot read your conversations casually, but logs exist and can be accessed in an investigation. That is not a reason to avoid it. It is a reason not to treat any account, personal or institutional, as a diary.

Checking your settings, once

Spend five minutes doing this properly:

1. Find the data controls or privacy settings. Look for a switch about improving the model with your data. Decide and set it.
2. Look for chat history retention and whether you can delete conversations. Deleting usually removes them from your view; retention for a period on the provider's side is normal.

3. Look at any memory feature and read what it has stored about you. People are frequently surprised.
4. Check whether your institution provides an account, and what its terms say.

Do it once, per tool, and then stop thinking about it.

The test before you paste

> *If this ended up in a public archive with my name on it, who would be harmed?*

If the answer is *nobody* — a lecture slide, a textbook page, your own essay draft — paste generously; the grounding lesson is right that supplied text beats remembered text. If the answer is *somebody who did not choose this*, it does not go in, and that holds regardless of what the settings say, because settings are a promise about handling, not a guarantee about outcomes.

BEFORE YOU MOVE ON

A nursing student removes the patient's name and asks a model about a case from placement, keeping the rare diagnosis, the town, and the admission month. Why is this still a problem?

A. Because clinical questions are outside what a general model is trained to answer accurately.
B. Because the free tier may train on the input, whereas a paid account would not.
C. Because the combination of rare condition, small location and date can identify the person even without a name.
D. Because placement material belongs to the hospital, so sharing it is a copyright matter that requires the trust's permission first.

17 · 8 min

Diagnosing a bad answer

THE ONE THING TO KEEP

Bad answers have five distinguishable causes — underspecified question, missing material, genuine error, degraded conversation, wrong tool — and each has its own fix, so diagnose before you reword.

Whose fault was it

When an answer is unusable, students do one of two things: blame the tool and give up, or blame themselves and rewrite the prompt at random. Neither is a diagnosis. There are five causes, they have different fixes, and you can usually tell them apart in about twenty seconds.

One: the question was underspecified

Symptoms. The answer is generic, starts at the beginning, covers the topic rather than the task, or is a balanced overview when you wanted a position.

Fix. The four elements: level, task, constraints, form. This is the most common cause by a distance, and the most common single omission is level.

Two: the model lacked the material

Symptoms. Vague where you needed specifics. Wrong about your course's notation, a local regulation, this year's syllabus, a niche paper, anything after its training cutoff.

Fix. Supply it. Paste the notes, the handbook, the paper. If you cannot supply it because it is not in your possession, this is not a prompting problem and no rewording will help — go and find the source.

Three: the model had it and got it wrong anyway

Symptoms. A confident specific that turns out to be false. An invented reference. A subtly wrong step in a derivation that otherwise looks fine.

Fix. Verification, not rewording. Ask again cold in a fresh chat; check the specific against a real source; run the computation in a deterministic tool. If it is wrong in the same way every time, the error is well-represented in training data — often a common misconception — and that is worth understanding rather than working around.

Four: the conversation went bad

Symptoms. It contradicts something it said earlier. It reintroduces an approach you rejected. It agrees with everything you say. It has forgotten a constraint you set.

Fix. New chat, carrying a short summary. Not more argument.

Five: the task is not one this tool does

Symptoms. Counting, exact arithmetic, alphabetising, checking whether something exists, telling you what happened last week, remembering your previous conversation.

Fix. A different tool, per Module 1. No prompt fixes a category error.

A worked diagnosis

A student asks for help with a statistics coursework and gets an answer recommending a t-test. The data is ordinal, from a questionnaire, with unequal group sizes.

- Was the question underspecified? Yes — they said *compare two groups* and never said the outcome was a five-point Likert item. **Cause one.**
- Did the model lack material? Partly — it never saw the data or the assignment brief. **Cause two.**

- Was it wrong anyway? Not exactly. Given what it was told, a t-test is a defensible default.

The fix is not a cleverer prompt. It is: state the measurement level, the design, the sample sizes and the actual question, and paste the brief. Then ask *which tests are defensible here and what does each assume*, and choose. Then run it in jamovi or JASP, which are free.

Notice that the diagnosis also taught the student something: they had not registered that measurement level determines the test. That is usually true. A bad answer often marks the exact place where your own understanding of the problem is thin, which is why diagnosing it beats rewording it.

The cause students most resist naming

Cause five — wrong tool — is the one people argue with. It feels like giving up. It is not; it is the same judgement a lab technician makes when they reach for a balance instead of a ruler. If the job is exact arithmetic, existence-checking, counting, or knowing what happened last Tuesday, no phrasing rescues it. The most experienced users of these tools are conspicuously quick to close the window and open something else, and that speed is most of what makes them look skilled.

The twenty-second routine

When an answer is bad, ask in this order:

1. Did I tell it who I am and what job I want done? *(one)*
2. Could it possibly have known this? *(two, five)*
3. Is the wrong part a specific — a name, number, date, reference? *(three)*
4. How long is this conversation? *(four)*

One pass through those four questions puts almost every failure in the right box, and each box has one fix. The point is not to become good at prompting. It is to stop wasting time rewording a question that was never the problem.

BEFORE YOU MOVE ON

A student's prompt is well specified and the source is pasted in, but the model gives the same wrong definition in three separate fresh conversations. What does the repetition indicate?

A. That the prompt still needs refining, since a consistent failure across attempts points to a consistent flaw in the instruction.
B. That the pasted source was too short for the model to ground on, so it fell back on memory each time.
C. That the model is retrieving a cached answer, so a different provider's model would give the correct version.
D. That the error is well represented in the training data, most likely a common misconception worth understanding.

18 · 7 min

Prompts worth keeping

THE ONE THING TO KEEP

Keep six to ten prompts you actually reuse in one plain file, and note that the clauses doing the real work are the restrictions — one at a time, do not re-write, do not help me.

A small file beats a large library

There is a market in prompt packs. Almost all of it is filler. What actually helps is a plain text file with six to ten prompts you have personally found useful, kept where you can copy from it in two seconds. Obsidian, Joplin, a note on your phone, a `prompts.md` file — the tool does not matter.

What follows are the ones worth starting that file with. They are written to be pasted and edited, not admired.

The quiz

> *Here is the material, between the markers.*
>
> *---*
>
> *[paste]*
>
> *---*
>
> *Ask me one question at a time on this material. Do not give me the answer with the question. Wait for my answer, then tell me whether it is right, what I missed, and ask the next one. Make roughly a third of the questions require applying the idea to a case not in the text. Keep going until I say stop, then tell me which topics I was weakest on.*

The *one at a time* and *wait* clauses are what make this work. Without them you get a list of ten questions and ten answers, which is reading, not retrieval.

The explain-back

> *I am going to explain [topic] to you from memory. Do not help me and do not fill gaps. When I finish, tell me: what was vague, what was actually wrong, what I omitted entirely, and what I clearly do not understand as opposed to merely mis-stated.*

You do the production; the model does the marking. This inverts the usual arrangement and is worth more than any explanation it could give you.

The marker

> *Here is the assessment brief and mark scheme, and here is my draft.*
>
> *---*
>
> *[brief and mark scheme]*
>
> *---*

> *[draft]*
>
> *---*
>
> *Act as the marker. For each criterion, say which band this currently sits in and quote the sentence that decides it. Then give the three changes that would move the most marks, in order. Do not rewrite anything and do not be encouraging.*

Do not rewrite anything is the essential clause. Without it you get an improved draft that is no longer yours.

The hint ladder

> *I am stuck on this problem. Give me hints one at a time, starting with the smallest possible nudge — a question about what I should be looking at, not a step. Only give the next hint when I ask. Never give the full solution unless I type SOLVE.*

The sceptic

> *Below is an argument. Act as a hostile examiner. Identify the weakest step, the strongest counterargument, and any claim that is doing more work than the evidence supports. Do not soften anything and do not list strengths.*

Use this in a fresh chat with the argument pasted in, not at the end of the thread that produced it.

The level ladder

> *Explain [topic] three times: first as you would to someone a level below me, then at my level using standard terminology, then telling me what the version at level two simplifies or hides and what a specialist would object to.*

The reading foothold

> *Here is a paragraph I cannot follow, and here is the sentence before and after it. Restate it in plain language without adding anything that is not there, then list the assumed background knowledge I appear to be missing. I will go back and read the original.*

Standing instructions

Anything you would paste every time belongs in the tool's custom-instructions field instead. Level, spelling convention, tone, and your standing preference for directness. Then your saved prompts can stay short.

Where a saved prompt goes wrong

A saved prompt is a fixed instruction meeting a changing situation, so it fails in one predictable way: it keeps working after it should have stopped. The quiz prompt written when you knew nothing about a topic will still be asking recall questions in week nine, when you need application. The marker prompt aimed at a 1,500-word essay will happily be applied to a lab report with entirely different criteria.

So read the prompt before you paste it, every time, and change the one clause that no longer fits. That takes five seconds and is the difference between a tool and a ritual.

The maintenance rule

Delete prompts you have not used in a month. A file of six live prompts gets used; a file of forty becomes a thing you scroll past. The best test of a prompt is whether you reached for it unprompted, and the best source of new ones is noticing the sentence you just typed that made an answer much better, and saving that.

BEFORE YOU MOVE ON

In the quiz prompt, why does "ask me one question at a time and wait for my answer" matter so much?

A. It reduces token use, so a free tier's allowance stretches further across a revision session.

B. Without it you receive questions and answers together, which is reading rather than retrieval.

C. It keeps the conversation short enough to avoid drift and truncation over a long revision session.

D. It lets the model adapt each question's difficulty to your previous answer, which is what makes the practice effective.

CASE STUDY · MODULE 2

Nine days, and one thousand nine hundred and fifty rupees

A second-year economics student in Guwahati, nine days from a macroeconomics paper, deciding whether the free tool is the thing that is wrong.

Sanchita Barua had a specific and repeated complaint about the free assistant she had been using since first year. It kept answering her course's questions in somebody else's course's terms.

Her department teaches the IS-LM treatment from one particular textbook, with its own notation and its own ordering of the derivation. The model gave her a general treatment, competent and clear, with different symbols and a different route to the same place. When she compared it against her lecture notes she could not always tell whether she was looking at a different convention or an error, which is a bad position to be in nine days before a paper.

She concluded that the free tier was the problem. A month of the paid tier was one thousand nine hundred and fifty rupees. Her allowance for data, mess extras and printing was about two thousand one hundred a month. She had decided to buy it, and had worked out which of the three she would cut.

A third-year in the next room suggested a test before spending, which cost an evening. They took five questions off her own past paper and ran each one twice. Round one was the bare question typed into the free tool as she would normally have typed it. Round two was the same question with three photographed lecture slides pasted in between markers, an instruction to use only what was in that text and to say what was missing rather than filling it in, and four things stated: that she was in the second year, that she wanted the derivation in her course's notation, that it should be under three hundred words, and that it should end with the step she was most likely to get wrong.

On three of the five, the grounded answers came back in her course's terms and were usable. On a fourth, the model used her notation and then reverted to the general form halfway down, which she could see at once because she

had the slide in front of her — so the check after grounding was not does this exist but does the passage actually say that, and it took ten seconds.

The fifth question was different. It turned on a specific figure from the Reserve Bank's policy corridor in 2019. Grounding could not help, because she had no source containing it to paste. She asked three times in three fresh chats and got three confident numbers, two of them different. No amount of rewording was going to fix that, and she eventually found the figure in a published policy statement in four minutes.

There was a real cost on the other side of this. Photographing forty slides, running them through free OCR on her phone and correcting the mangled subscripts took two of her nine evenings. She had planned to spend those evenings on past papers.

She also made one decision that week she describes as the only one she got plainly wrong. A classmate asked her to paste his unsubmitted answer into her chat for a check, and she did it without thinking about it for even a second. He had not agreed to a company holding his work, and she had not asked him. Saying no would have cost one mildly awkward sentence.

STOP HERE

Before turning the page: what would you have done, and what would it have cost if you were wrong? Write it down. The point of the next section is to compare.

WHAT HAPPENED

She did not buy the subscription. Nine days later she scored sixty-eight on the paper, against fifty-four on the previous one, and she is careful to say that she also spent nine days revising and that no part of this is a controlled experiment.

The finding she took from it is narrower and more useful than the mark. Her complaint had a cause and the cause was not model quality. The model lacked the material — her course's notation lived in her notes and nowhere in its training in that form — and supplying the material fixed it, at a cost of two

evenings. Buying a stronger model would have bought a better general treatment of something she did not need a general treatment of. The test also showed her where a stronger model might genuinely have helped, on the one hard derivation where both versions were thin, and she never tested that, so the case does not show that grounding beats a better model in general. It shows that she had diagnosed the wrong cause and was about to spend a month's printing budget on it.

Two habits survived the paper. She now starts a fresh chat per topic, after a forty-message thread reverted twice to a definition she had already corrected in it. And she keeps her lecture slides as text in one file, because the expensive part was never the pasting.

Her original failure has one of the five causes. Which, and why did the diagnosis 'the free model is not good enough' point her at the wrong fix?

It is cause two: the model lacked the material. The symptoms fit exactly — vague or divergent on her course's notation, on a local convention, on the specific treatment her department uses — and the fix for cause two is to supply the material, not to reword and not to upgrade. She read the symptom as cause three, the model having it and getting it wrong, which is the one that would have justified paying for a better one. The distinction matters because the two causes have opposite fixes and similar surfaces: an answer that is confidently in the wrong convention and an answer that is confidently false look much the same from the outside. The twenty-second routine separates them by asking whether the tool could possibly have known this, and her lecturer's notation is something it could not.

On the Reserve Bank figure, grounding did not help at all. Which cause was that, and what was the correct move?

It was cause five, the wrong tool for the task, sitting on top of cause two. A specific policy number from a particular year is exactly the weakly-represented category: stated in a few places, reproduced in a rigid numeric form, and generated rather than recalled. There is nothing to paste because she does not hold the source, and no prompt rescues a category error. The correct move is the one she eventually made — go to the primary record, which for a central bank statement is free and online and took four minutes. The three fresh chats were not wasted, though. Two

disagreeing answers to the same cold question are strong evidence that you need a real source, which is the cheapest reliability test available.

She pasted a classmate's draft. Why is that different in kind from pasting her own lecture slides, and what does a privacy setting not fix?

She can consent to a service handling her own material. She cannot consent on somebody else's behalf, and an unsubmitted draft is his work and his data. That distinction does not move with the settings, because a setting is a promise about handling rather than a guarantee about outcomes: companies change, are acquired, are breached and are subject to legal process, and a training toggle set today says nothing about any of those. The workable test is to ask who would be harmed if this appeared in a public archive with a name on it. For a lecture slide the answer is nobody, and she should paste generously. For a person who did not choose it, the answer is somebody, and it does not go in whatever the settings say.

MODULE 2 · TEST

Check what stuck

6 questions, one right answer each. This one is for you, not for a record: answers and the reason behind each are at the back. If two questions from the same module go wrong, re-read that module before the exam rather than after.

1 An answer comes back pitched far below where you are, starting from definitions you already know. Which of the four elements did you omit?

A. Level and prior knowledge
B. The specific task to be done
C. Constraints on length and omissions
D. The form the answer should take

2 You paste the real passage and ask the model to use only that text. What check does grounding replace, and with what?

A. Nothing now needs checking, since the source was supplied
B. You check the summary against the passage in front of you
C. You read the model's confidence rating before the passage
D. You ask the model to confirm it used only the supplied text

3 The rephrase-and-repeat reliability check must be run in a new chat. Why does continuing the same one not work?

A. A new chat resets the model's temperature to a lower value
B. Fresh chats are served by a more recent version of the model
C. The daily quota resets, so the second answer costs nothing
D. Everything in the context conditions the next answer

4 Forty messages in, an assistant contradicts itself, reintroduces an approach you rejected and agrees with everything you say. What is the repair?

A. Argue with it until it withdraws the claim it keeps repeating
B. Ask it to re-read the earlier messages before each new reply
C. Start a fresh chat carrying a short summary you have checked
D. Raise the word limit so the whole conversation fits the window

5 Where does the line sit on what may go into a chat window?

A. You can consent for your own data, never for somebody else's
B. Anything is safe once the training toggle has been switched off
C. Only material that is already published may be pasted at all
D. Taking the names out of a case makes it safe to paste anywhere

6 Asking for step-by-step working gives you two real things. Which is not one of them?

A. Proof that the steps shown are what produced the answer
B. Better accuracy, because generating steps helps on multi-step problems
C. Six small things you can verify instead of one you cannot
D. A place to apply the units, magnitude and joint checks

MODULE 3

Study that raises the difficulty

The techniques with the strongest evidence behind them all feel worse than the ones that do not work. Hint ladders, retrieval, spacing, interleaving and self-explanation, rebuilt around a tool that will otherwise hand you the answer.

> BY THE END OF THIS MODULE YOU CAN
>
> Run a study session in which the assistant raises the difficulty rather than lowering it — a hint ladder instead of a solution, retrieval before review, spaced and interleaved practice, explanation produced by you — and demonstrate afterwards what you can do with the window closed

19 · 8 min

Getting Unstuck, Not Getting Finished

> THE ONE THING TO KEEP
>
> Ask for the smallest hint that lets you continue, then finish it yourself.

Four kinds of stuck

"I'm stuck" is four different situations, and each wants a different amount of help. Naming which one you are in is most of the skill.

1. **I don't understand what is being asked.** Cheap to fix, no learning cost. Get it clarified immediately.
2. **I'm missing a fact or a formula.** Also cheap. Knowing that the integral of 1/x is ln|x| is lookup, not insight.

3. **I know the method but the execution has gone wrong.** This is where the learning is. Guard it.
4. **I have an answer and it's wrong and I can't see why.** This is where the *most* learning is, and it is the one people give away fastest.

Cases 1 and 2 — just ask. Nobody builds character by not knowing what a word means. Cases 3 and 4 are the ones where the phrasing of your request decides whether you come out of it knowing more.

The minimum viable hint

Models are trained to be helpful, and left alone they will hand you the whole thing. You have to say otherwise, explicitly, every time. These work:

```
Do not give me the answer. Ask me one question that would
get me unstuck.
```

```
Give me the smallest hint that lets me continue. If that
isn't enough I'll ask again.
```

```
Here is my working through line 6. Is line 6 correct?
Answer yes or no. Nothing else.
```

```
I think the problem is in my choice of method, not my
arithmetic. Am I right about that? Don't say which method.
```

That last one is worth practising. You are asking it to confirm the *shape* of your error while you keep the work of locating it. It is the highest-value question in this lesson.

Most chat tools let you set a standing instruction — custom instructions, a saved system prompt, a project. Put the hint rule there once and stop retyping it:

```
When I ask about schoolwork, default to hints and questions.
Give full solutions only when I say "full solution".
```

Debugging is the clearest case

Code makes this concrete because the error message is right there.

```
TypeError: 'NoneType' object is not subscriptable
```

The fast path is to paste your file and the error and get corrected code back. You will ship in ninety seconds and hit the same error next week, because `NoneType` errors are a category, not an incident.

The slower path:

```
Here's the traceback. Don't look at my code yet.
What does this error class usually mean, and what are the
two or three most common causes?
```

Then you go and find yours. Third time you meet that error, you will not need to ask at all. That is the whole return on the extra four minutes.

Struggle is not the ingredient

Here is where a lot of study advice goes wrong, and it is worth being precise.

The research on productive failure — Manu Kapur's work, and the desirable-difficulties literature generally — is about struggle *followed by instruction*. Attempt, fail, then get taught. Struggle alone, with no resolution, is not a learning condition. It is just an unpleasant afternoon.

So timebox it. A workable rule: **if you have made no forward progress in six or seven minutes, the struggle is done producing value.** Not the six minutes when you were slowly assembling something. The six minutes of rereading the same line.

At that point, get the smallest hint and carry on. You have banked the benefit. Sitting there for another twenty minutes out of pride costs you the next topic.

Close the loop

One step that people skip, and it converts the whole thing.

After the hint got you moving and you finished the problem — **do it again from a blank page, later that day.** No hint, no notes. Two minutes if it went in. If it did not go in, that is worth knowing now, cheaply, rather than in an exam hall.

And keep a running list of what you needed hints for. Not to feel bad about. That list is a map of exactly where your understanding is thin, and it is far more accurate than any feeling you have about which topics you are weak on. When revision starts, you will not have to guess.

BEFORE YOU MOVE ON

You have been on the same integral for 25 minutes and have made no progress since minute six. Which move is most defensible?

A. Keep going unaided. Struggle is the mechanism that builds the memory, and ending it early throws away the benefit you have been earning.

B. Ask for the full worked solution. You have put in honest effort and the remaining benefit of holding out is small.

C. Move on to the next question, since your time is better spent on problems you can actually complete.

D. Ask for the smallest hint that unblocks you, finish the problem yourself, then redo it from blank paper later that day.

20 · 7 min

Make It Quiz You

THE ONE THING TO KEEP

A model that asks you questions is worth more than one that answers them.

The single highest-return change

Most students use AI as an answering machine. The version that actually moves exam scores is the one where it asks and you answer.

The reason is one of the best-supported findings in the study of memory. Roediger and Karpicke (2006) had one group reread a passage four times, and another read it once then try to recall it repeatedly. Tested five minutes later, the rereaders won. Tested a week later, the recall group remembered far more — the gap was large, roughly 60% against 40% recall in one of their experiments. And the rereaders had confidently predicted the opposite.

This is called the testing effect, and "testing" is a bad name for it, because it makes it sound like measurement. It is not measurement. **The act of pulling something out of your head changes how available it is next time.** A quiz is not a check on studying. It is studying.

AI is unusually good at this job, because the limiting factor was always that somebody has to write the questions.

The prompt that works

The key details are: one at a time, wait, no hints, and grade against what you actually said.

```
Here are my lecture notes on the Krebs cycle.

Ask me one question at a time. Wait for my answer before
the next one. Do not give hints unless I ask. After each
answer, tell me only whether I'm right and what I left out.
Don't teach. Ten questions.
```

What you must watch for: after two or three rounds, the model will start volunteering explanations, and your session will quietly turn back into reading. Re-anchor with `Stop explaining. Next question.`

Some variants worth having:

```
Ask me the question my exam is most likely to ask that
my notes would not prepare me for.
```

```
I'm going to explain photosynthesis to you as if you're 12.
Interrupt at the first point where what I say stops being
true, and don't correct it — just tell me where you stopped.
```

That second one is the Feynman check, and the stopping point is the whole value. It locates the exact sentence where your understanding runs out, which is almost never where you would have guessed.

Writing your own questions beats having them written

If you have time for only one thing, do this instead. Read a chapter, close it, and write eight exam questions on it. Then hand your questions to the model and ask which important ideas you failed to ask about.

This works for two reasons. Generating something yourself makes it stick harder than reading the same thing — the generation effect. And to write a good question you have to know what the chapter's load-bearing ideas were, which is a stiffer test than answering.

The model's job here is to find your blind spots, not to do the thinking.

Two honest warnings

It will get facts wrong, and you cannot audit it. You are asking a machine to quiz you on material you do not yet know, which is exactly the situation where you cannot catch its errors. So: quiz from *your own* notes and set texts, pasted in, rather than from the model's general knowledge. And when it corrects you on something and you feel surprised, check the textbook before you update. Surprise is the signal.

This matters more the further you get from big, well-covered subjects. Introductory biology in English: mostly reliable. Your national curriculum's specific treatment of a local legal code, in a language with less material online: much less so.

A quiz you pass immediately after reading proves almost nothing. The material is still sitting in working memory. Real retrieval practice happens the next day, or three days later. Same questions, cold. That is when the answer either comes or does not, and that is the answer that predicts the exam.

What a week looks like

- **Day 0:** read the chapter. Write eight questions from memory.
- **Day 1:** answer your own questions cold. Get the model to mark you against the source text.
- **Day 3:** only the ones you missed.
- **Day 8:** all eight again, plus whatever came up in the lecture since.

Four short sessions, maybe eighty minutes total. It will feel less productive than eighty minutes of rereading with a highlighter. That feeling is the point — it is the difficulty doing the work.

BEFORE YOU MOVE ON

You reread a chapter three times and feel confident. A friend reads it once, then answers twenty questions from memory, gets many wrong, and feels shaky. Who is likely to know more a week later, and why?

A. The friend, because attempting recall changes what you can retrieve later; your confidence was fluency with the page rather than access to the content.
B. You, because you covered the material three times and repeated exposure is what builds memory, while your friend saw it once.
C. Neither reliably, because the friend's high error rate means they encoded wrong answers, and wrong answers are worse than no answers.
D. The friend, but only because twenty questions happened to cover more of the chapter than three readings did.

21 · 8 min

Building a hint ladder

THE ONE THING TO KEEP

Take the smallest hint that lets you continue, and when checking work ask only for the number of the first wrong line — the gap between locating an error and being told the fix is where the learning is.

The size of the hint decides what you learn

When you are stuck, there is a whole range of possible help between nothing and the answer, and almost all of the learning lives in the small end. A hint ladder is that range made explicit, so you take the smallest rung that gets you moving instead of jumping to the top.

A five-rung ladder for a problem in any subject:

1. **Attention.** *Which part of the question have you not used yet?* No content at all, just a direction to look.
2. **Category.** *This is a conservation-of-energy problem, not a kinematics one.* Names the class without touching the instance.
3. **First move.** *Start by writing the energy at the top and the energy at the bottom.* One step, no working.
4. **Structure.** The full outline with the arithmetic left out.
5. **Solution.** Everything.

The rule is simple: take rung one, work for two minutes, take rung two only if still stuck. Most problems break at rung two or three. The rungs above four should feel like a defeat you have earned, not a default.

Take the smallest rung that gets you moving

Rung one: attention	Which part of the question have you not used yet? No content at all.
Rung two: category	This is conservation of energy, not kinematics. Names the class, not the instance.
Rung three: the first move	Write the energy at the top and at the bottom. One step, no working.
Rung four: structure	The full outline with the arithmetic left out.
Rung five: the solution	Everything. A defeat you have earned rather than a default.

Most problems break at rung two or three. Attempting to produce an answer before being shown it improves retention of the correct answer even when the attempt fails, which is why a ladder is a machine for getting as many attempts as possible out of one problem and a solution is a machine for getting none. Expect the model to drift upward; that is what its tuning means by helpful.

Why the small rung is worth more

The effect underneath this is **generation**: attempting to produce an answer before being shown it improves retention of the correct answer, even when the attempt fails. It has been replicated across vocabulary, concepts and problem solving. Failed generation followed by feedback beats being told directly, which is counterintuitive enough that most students will not believe it until they run the notebook experiment from Module 1 on themselves.

A hint ladder is a machine for extracting as many generation attempts as possible from one problem. A solution is a machine for extracting none.

The prompt, and the clauses that matter

> *I am stuck on the problem below. Give me hints one at a time, starting with the smallest possible nudge — a question about what to look at, not a step. Wait for me to try. Only give the next hint when I ask. Never give the full solution unless I type SOLVE. If my attempt is wrong, tell me which line is wrong without telling me what it should be.*

The last sentence is the one people leave out and it is the most valuable. *Which line is wrong* points you at the error; *what it should be* removes the work. The gap between those two is where the learning happens.

Locating the error without revealing it

A related move for when you have an answer that does not match:

> *Here is my working. Tell me only the number of the first line that contains an error. Nothing else.*

This is astonishingly effective on long derivations, and it is the closest thing to a good human tutor's behaviour that these tools do naturally. You get the location, you do the diagnosis. If you still cannot see it after a genuine attempt, escalate: *what kind of error is on line 4 — algebraic, conceptual, or an unjustified assumption?* Still no content, more direction.

When to abandon the ladder

The ladder is for problems within reach. It is the wrong tool in three cases:

- **You are missing the prerequisite.** If you do not know what a partial derivative is, no hint about the problem helps. Recognise this and go back a level rather than grinding. The tell is that rung two makes no sense to you.

- **Time has run out.** At eleven at night with a deadline, take the solution, and then — this is the part that matters — mark the problem and redo it from scratch two days later. A solution taken under pressure is not a failure unless you never come back to it.
- **The problem is broken.** Sometimes the question has a typo or is genuinely ambiguous. Two rungs of hint that make no sense is a reason to check with a human, not to assume you are stupid.

Building your own ladders for others

If you tutor, mentor, or work in a study group, writing hint ladders is the single most useful thing you can prepare. Take the six problems everybody gets stuck on and write five rungs for each. It takes an hour and it is more useful than any amount of explaining, because it lets the person you are helping keep doing the work.

It is also, incidentally, one of the best ways to find out whether you understand a topic. Writing rung two — the one that names the category without touching the instance — requires you to know what the problem is really about.

The honest limitation

A model will drift up the ladder if you let it. Assistants are tuned to be helpful, and *helpful* in that tuning means complete. Expect it to sneak the answer into rung two, and expect to have to say *that was too much, go back to a smaller hint*. This is not a failure of your prompt. It is a standing tendency you manage every session, which is why the SOLVE keyword exists — it gives the model an explicit thing to wait for.

BEFORE YOU MOVE ON

A student's hint prompt says "tell me which line is wrong" but they omit "without telling me what it should be". What is lost?

A. Nothing significant, since knowing which line is wrong already tells you where to look and the correction only saves time.

B. Accuracy, because a model asked to correct a line is more likely to hallucinate a fix than to identify a fault.

C. The generation attempt: locating an error is direction, diagnosing it is the work.

D. The record of the attempt, since a corrected line overwrites the student's own working and weakens their process evidence.

22 · 9 min

Retrieval before review

THE ONE THING TO KEEP

Write from a blank page before you open anything, because the gap that matters is what you were confident about and had wrong, and no fluent explanation will ever show you that.

The blank page comes first

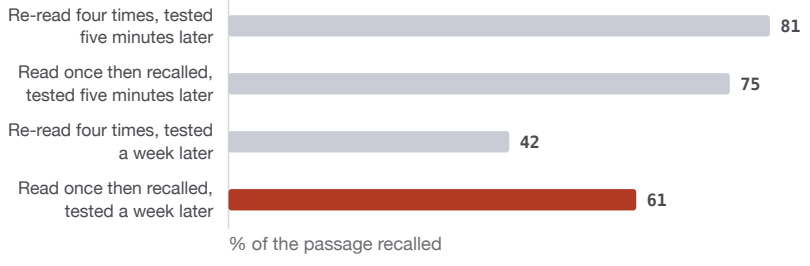
The most robust finding in the study-methods literature is that trying to recall something strengthens memory for it more than reading it again does. It is called the **testing effect** or retrieval practice, it has been replicated for decades across ages and subjects, and it is close to free.

The standard demonstration: two groups study a passage. One re-reads it repeatedly. One reads it once and then repeatedly tries to write down what it said. On a test five minutes later, the re-readers do slightly better. On a test a

week later — which is the one that resembles an exam — the retrieval group does substantially better, and the gap is not small.

The part that matters for anyone using an AI assistant: the re-readers were also more confident. Retrieval practice makes you feel less prepared and leaves you more prepared. Every incentive in the moment points the wrong way.

The method that felt worse is the one that held



Roediger and Karpicke's demonstration in the shape it is usually reported. The re-readers won the immediate test, lost the one a week later, and had confidently predicted the opposite. Retrieval practice makes you feel less prepared and leaves you more prepared, so every incentive in the moment points the wrong way — which is exactly why an assistant that explains fluently is so easy to mistake for study.

The order that makes an AI session work

Most students open the model and ask it to explain. That is review before retrieval, and it is backwards. The order that works:

1. **Blank page.** Write everything you can about the topic from memory. Five minutes, no notes, no window open. It will feel thin. That is information.
2. **Then open the source.** Your notes, the textbook, or the model with your notes pasted in.
3. **Mark the gaps.** What did you miss entirely? What did you have half right? What did you have confidently wrong — the most important category.
4. **Close everything and write again.** Not later. Now.
5. **Then quiz.** Use the model, with the material pasted, one question at a time.

Step one is the one people skip and the one carrying most of the effect.

Why confidently wrong is the valuable category

Things you knew you did not know are cheap to fix; you already knew to look them up. Things you were sure about and had wrong are what cost marks, because you will not check them. The blank page is the only cheap way to find them, since a fluent explanation you read along with never gives you the chance to be wrong.

When you find one, write it down separately with the correction. That list — your personal confidently-wrong list — is the most valuable revision document you will own by the end of a term, and it takes no extra time to build.

Making the model do retrieval properly

The quiz prompt from Module 2 works because of two clauses. Add a third for depth:

> *Make roughly a third of your questions require applying the idea to a case that is not in the material.*

Recall questions test whether the words are available. Application questions test whether the idea is. Exams mostly ask the second kind, and models default to the first because it is easier to generate from a text.

And insist on this:

> *Do not accept a vague answer. If I give one, ask me to be specific before you tell me whether I am right.*

Without it, assistants mark generously. A model that says *yes, roughly* to a half-answer has taught you that a half-answer is enough, which is exactly the wrong lesson to carry into an exam hall.

Free tools that do this well

Anki is free on desktop, Android and the web, and is the best implementation of spaced retrieval that exists. iOS is the exception and is paid; **AnkiDroid** and the desktop version are free and sync with each other.

You can use a model to draft cards, and this is a genuinely good use — *turn this passage into cloze-deletion cards, one fact each, no card longer than fifteen words* — with one caution. Cards you wrote yourself are better remembered than cards you were given, for the same generation reason as everything else in this lesson. A reasonable compromise: write your own cards for the material that matters, generate cards for high-volume factual material like vocabulary, anatomy or dates, and check every generated card against the source before it enters the deck. A wrong card drilled for six months is expensive.

What retrieval does not do

It is not a substitute for first learning the material. You cannot retrieve what was never encoded, and a blank page on a topic you have not yet studied produces nothing but frustration. The sequence is: encounter it properly once, then move to retrieval immediately and stay there. Most students invert that too, spending 80 per cent of their time on encountering and 20 on retrieval, when something closer to the reverse is right.

BEFORE YOU MOVE ON

Two groups study the same passage for the same total time. Group A re-reads it four times; group B reads it once and then attempts to write it from memory three times. Group A scores higher on a test five minutes later. What should you conclude?

A. That the retrieval advantage is real but small, so the choice between methods matters less than total study time.

B. That re-reading suits factual passages while retrieval suits problem solving, so the method should match the material.

C. That group B's attempts were probably poorly executed, since correctly performed retrieval practice should win at any delay.

D. Nothing much, because the immediate test is where re-reading looks best and the delayed test is the one that matters.

23 · 8 min

Spacing and interleaving

THE ONE THING TO KEEP

Forgetting a little between sessions is the ingredient that makes spacing work, and mixing problem types without labelling them is what trains the recognition an exam actually tests.

The same hours, arranged differently

Take six hours of revision for one topic. Spend them in one Sunday block and you will remember substantially less in three weeks than if you spend them as six one-hour sessions across three weeks. Nothing else changes. This is the **spacing effect**, and it is one of the oldest results in psychology, going back to Ebbinghaus in the 1880s and confirmed continuously since.

The mechanism is worth knowing because it makes the advice stick. Each time you retrieve something after partly forgetting it, the retrieval is effortful, and effortful retrieval strengthens the memory more. Study it again immediately and there is nothing to retrieve — it is still sitting in mind — so the repetition does almost no work. **Forgetting a little is the ingredient.**

That is why massed study feels so much better. Everything is fresh, nothing is hard, and the smoothness reads as competence. It is the fluency trap in a different costume.

Intervals that are good enough

People get lost trying to optimise the schedule. The gains from perfect intervals over roughly sensible ones are small. Roughly sensible:

- Same day, a few hours later.
- Two days later.
- A week later.

- Two to three weeks later.
- A month before the exam, and then once in the final week.

A rough guide that works: if the exam is *n* weeks away, the gaps between reviews can be around *n* divided by five. Nothing turns on the exact number. What turns on it is that a gap exists at all.

Anki implements this automatically and adjusts per card by how you rate the difficulty. If you do not want a system, a paper calendar with topic names on dates is entirely sufficient and takes ten minutes to build.

Interleaving: the harder one

Blocked practice is twenty problems of type A, then twenty of type B. Interleaved practice mixes them: A, C, B, A, C, B.

Interleaving produces worse performance during practice and better performance on a later test. In the classic demonstrations with mathematics problems, students practising in blocks outperform the interleaved group during the session by a wide margin and are then beaten by an equally wide margin on the test a week later.

The reason is specific and useful. In a block, you already know which method to use — the heading told you. On an exam paper, nobody tells you. Choosing the method is a separate skill from executing it, and blocked practice never exercises it. Interleaving is how you learn to recognise which kind of problem you are looking at, which is precisely what students who *knew the material* and still did badly were missing.

Blocked practice wins the session and loses the test

	During the session	A week later
Blocked	**89%** feels efficient	**38%** the heading told you
Interleaved	**63%** feels worse	**74%** you chose the method

The shape of the classic mathematics demonstrations, with illustrative percentages. In a block you already know which method to use, because the heading told you. An exam paper does not tell you, and choosing the method is a separate skill from executing it. Interleaving is how that skill gets trained, and it feels worse for the whole of the session in which it is working.

Getting a model to interleave

Models default to blocking, because a set of similar questions is the natural output of a prompt about one topic. You have to ask:

> *Here are the six topics from this module. Give me one question at a time, drawn at random from any of them, without telling me which topic it is from. After my answer, tell me the topic and whether I was right.*

Without telling me which topic is the whole technique. That clause converts an exercise in execution into an exercise in recognition.

A harder variant for exam preparation:

> *Mix in occasional questions that look like topic A but are actually solved with topic B's method.*

Those are the ones that separate grades, and they are almost impossible to construct for yourself, because you know the answer.

What is not worth your time

Learning styles. The idea that you are a visual or auditory learner and should be taught accordingly has been tested repeatedly and does not hold up; matching instruction to a claimed style does not improve outcomes. Believe your preferences about what is pleasant, not about what works.

Highlighting. Consistently rated among the least effective techniques when studied. It feels like engagement and produces recognition, not recall. If you highlight, treat it as a way of marking what to later retrieve, never as the study itself.

Re-copying notes. Same problem. Copying is not encoding. Rewriting notes *from memory* is a different activity entirely and is excellent.

The scheduling conversation worth having

One good use of an assistant is planning:

> *I have five topics and three weeks. I can do 45 minutes a day, five days a week. Build a schedule that spaces each topic at increasing intervals and interleaves them within each session. Give it to me as a plain list by date.*

The model is good at this because it is a constraint-satisfaction problem with a known answer shape. Check the arithmetic — it is the kind of thing it gets slightly wrong — and then follow it, which is the hard part and is not the model's job.

BEFORE YOU MOVE ON

Why does interleaved practice beat blocked practice on a later test despite producing worse scores during the practice session itself?

A. Because a block tells you which method to use, so it never trains recognising the problem.

B. Because mixing topics increases total time on task, as switching forces you to re-read each question more carefully.

C. Because interleaving spaces each topic out, so the benefit is really the spacing effect under another name.

D. Because varied practice keeps attention higher, and sustained attention is the main determinant of what is retained from a session.

24 · 8 min

Explaining it back

THE ONE THING TO KEEP

Explaining from memory to a listener who refuses to help produces the learning; the most useful feedback to ask for is which parts you have memorised correctly without understanding.

Who is doing the talking

There is a decisive difference between asking a model to explain something and explaining it to the model. The first is reading. The second is producing, and producing is where learning happens.

The research name for the second is **self-explanation**: while studying, saying out loud how each step follows and why. Studies since the late 1980s find that students who spontaneously self-explain while working through examples learn considerably more from the same material than those who do not, and that prompting students to self-explain raises outcomes for everybody.

A language model is an unusually good self-explanation partner, for one boring reason: it will listen to eight hundred words of confused reasoning at eleven at night and respond specifically. No human tutor is available on those terms.

The prompt, and the discipline it needs

> *I am going to explain [topic] to you from memory, as if you know nothing about it. Do not help me. Do not fill in gaps. Do not tell me I am doing well. When I finish, tell me: which parts were vague, which were actually wrong, which I skipped, and which I appear to have memorised without understanding.*

Then write it. Properly, in full sentences, without opening your notes. Ten minutes.

The discipline: **do not look anything up while writing.** The moment you check, the exercise becomes copying. The gaps are the output. If you cannot continue, write *I do not know how the next part works* and carry on from the other side.

The distinction that makes the feedback useful

The fourth item in that prompt — *memorised without understanding* — is the one that earns its place. It asks the model to distinguish between two things that look identical on paper: a correct statement of a rule, and a grasp of why the rule holds.

The tell is usually that you can state the result and not the condition. *The mean of the sample estimates the population mean* — correct, and it does not tell anybody whether you know what randomisation is doing there. Asking the model to probe for that catches a very common state: fluent, correct, and hollow.

Follow it up with the question that settles it:

> *Ask me three questions that would distinguish someone who understands this from someone who has memorised the standard statement.*

Elaborative interrogation: the why chain

A close relative, with its own evidence base. For each fact, ask *why is that true?* and answer it. Then ask why of the answer. Three levels is usually enough to reach either bedrock or the edge of your understanding, and both are useful destinations.

> *Metals conduct electricity. Why? Delocalised electrons move freely. Why are they delocalised? The outer electrons are weakly held and the lattice of positive ions shares them. Why does that lattice form?*

Models are good at being the *why* in this chain and terrible at being the answer if you let them, because after two levels they will start producing plausible depth. Keep answering yourself and use the model to ask.

Explaining to a specified audience

Changing who you are explaining to changes what you have to know.

- **To a 12-year-old.** Forces you to drop jargon, which is where hidden reliance on terminology shows.
- **To a sceptical expert in a neighbouring field.** Forces you to justify rather than assert.
- **To someone who holds the opposite view.** Forces you to know what the disagreement is actually about.

Ask the model to play one of these and to stay in character, including asking the questions that audience would ask. It does this well, and the questions from the sceptical-expert version are usually the ones your marker would ask.

The version with no technology

All of this predates the tools and works without them. Explain it to a friend, a family member who does not know the subject, an empty room, or the voice recorder on your phone. The value is in the production, and the model is a convenience — a listener who is always awake and will tell you where you were vague. If your data has run out, talk to the wall. It works, minus the feedback.

Where it fails

Self-explanation works on material you have already met once. It does nothing on material you have not encountered — there is nothing to explain — and students sometimes conclude they are bad at it when they have simply skipped the reading. Encounter first, briefly. Then explain. Then read again to fill what the explanation exposed.

BEFORE YOU MOVE ON

In the explain-back prompt, why is "do not fill in gaps" a necessary instruction rather than an obvious one?

A. Because gaps in a student's account are usually caused by nerves rather than ignorance, and filling them masks a confidence problem.
B. Because assistants are tuned toward completeness and will supply the missing step, turning production back into reading.
C. Because a filled gap enters the conversation context and will bias every later question the model asks you.
D. Because the model's version of the missing step may be wrong, and the student has no way to tell.

25 · 8 min

Worked examples, and when to stop using them

THE ONE THING TO KEEP

Worked examples help beginners and hinder people who already have the method, so fade from full example to blank problem by removing steps from the end — and cap how many examples you read before moving on.

Beginners and experts need opposite things

When you are new to a type of problem, studying a fully worked example teaches you more per minute than attempting the problem yourself. This is well established, and it is the one place where being shown the answer is the right move.

The reason is capacity. Working memory holds only a few items at once. A beginner attempting an unfamiliar problem spends all of it on searching — what do I do first, is this the right formula, what was that rule — and has nothing

left for noticing the structure. A worked example removes the search, so attention goes to the pattern.

But this reverses. The same worked examples that help a beginner actively slow down someone who already knows the method: they now have to process an explanation of something they can already do, which costs capacity and gives nothing. This reversal has a name, the **expertise reversal effect**, and it is why study advice that ignores your current level is unreliable.

Fading, which is the actual technique

The right sequence is not *examples* or *problems*. It is a fade from one to the other:

1. **Full worked example.** Every step shown, with the reason for each.
2. **Completion problem, last step removed.** You do the final step.
3. **Last two steps removed.**
4. **Only the first step given.**
5. **Nothing given.**

Each stage is a small generation attempt at the point you are most ready for it. This structure — *backward fading*, because the removal starts from the end — outperforms both pure examples and pure problem solving in controlled comparisons.

Models are unusually good at generating this, and almost nobody asks them to:

> *Give me a worked example of this type of problem with the reasoning at every step. Then give me the same type of problem with the last step missing for me to complete. Then one with the last two missing. Then one with only the first step given. Then one with nothing. Wait for my answer at each stage before continuing.*

That is a complete, well-evidenced practice sequence, generated in one message.

Study the example properly, which most people do not

Reading a worked example passively gets you very little. The version that works:

- After each line, cover the rest and predict the next step before revealing it.
- Write, beside each step, *why* that step is legal. Not what it does — why it is allowed.
- At the end, close the example and write the sequence of moves from memory, in words, without numbers. This is the pattern you are actually trying to acquire.

That last one is the difference between learning to do *this* problem and learning to recognise this *kind* of problem.

Knowing when to stop

The signal that you have crossed out of the worked-example stage is that you can predict the next step correctly before revealing it, twice in a row. From that point, examples cost you time. Move to completion problems immediately, and to blank problems soon after.

Students overstay in this stage badly, because reading examples is comfortable and pleasant and feels like work. The number of worked solutions somebody has read is close to uncorrelated with what they can do; a solution read is a solution you can follow.

The trap specific to AI

A model will produce an unlimited supply of worked examples on demand. That makes it easy to spend an entire evening in stage one, which feels productive and is barely study at all. The fix is procedural: **decide in advance how many worked examples you will read before moving to completion problems.** Two is usually right; three if the material is genuinely new. Then move, whether or not you feel ready, because feeling ready is not a reliable signal here.

Checking that the example is right

One caution the classic research did not have to deal with. Worked examples in a textbook were checked. A generated one was not. In any subject where the steps matter, verify the example against a source before you learn from it — a textbook problem, a past paper with a mark scheme, a symbolic tool like SymPy or Maxima for the algebra. Learning a wrong worked example is worse than learning nothing, because you will apply it confidently.

The safest arrangement: take the worked example from your course materials, and use the model to generate the *fading sequence* of practice problems around it.

BEFORE YOU MOVE ON

A student who can already solve a class of problem is given detailed worked examples of it. Why does this measurably slow them down?

A. Because it reduces motivation, and material that feels too easy lowers the effort they put into the session.

B. Because worked examples always show one particular method, which interferes with the method the student had already learned.

C. Because reading a solution suppresses the retrieval attempt, and retrieval is the only mechanism by which practice strengthens memory.

D. Because processing an explanation of something already known consumes working memory and returns nothing.

26 · 9 min

A study session that works

THE ONE THING TO KEEP

Bookend every session with a blank page, cap the assisted block with a timer, and close with three lines saying what you can now do unaided and what you cannot.

Fifty minutes, arranged

Everything in this module assembles into one session structure. Here it is with the clock attached, because a technique without a shape does not get used.

0-5: blank page. Write what you remember about today's topic. No notes, no window. It will be thin. Keep it — you are going to compare against it at the end.

5-10: mark the gaps. Open your notes or the source. Three colours or three columns: missed entirely, half right, confidently wrong. The third column is the valuable one and goes on your running list.

10-25: work the gaps, hint-laddered. This is where the assistant comes in. Take one gap. Ask for the smallest hint. Work. Escalate only when stuck. If it is a problem type you cannot do at all, ask for one worked example, study it by predicting each step, then a completion problem.

25-30: break. Actually leave the desk. This is not optional filler; consolidation happens in the gaps and the next block is worse without it.

30-45: interleaved retrieval. The quiz prompt, mixing today's topic with two earlier ones, one question at a time, no topic labels, a third of the questions applied to new cases.

45-50: blank page again, and the log. Rewrite the topic from memory. Compare with your 0-5 version. Then write three lines: what you can now do unaided, what you still cannot, and what you will review in two days.

Fifty minutes, with the clock attached

0 to 5 min	Blank page. Write what you remember with nothing open. It will be thin, and the thinness is the information.
5 to 10 min	Mark the gaps in three columns: missed entirely, half right, confidently wrong. The third column is the valuable one.
10 to 25 min	Work the gaps on the hint ladder, smallest rung first. Timer on, because this is the block that swallows the session.
25 to 30 min	Break, away from the desk. Consolidation happens in the gaps and the next block is worse without it.
30 to 45 min	Interleaved retrieval. One question at a time, today mixed with two older topics, no labels.
45 to 50 min	Blank page again, then three lines: what you can now do unaided, what you cannot, what you review in two days.

Every block is one of the effects in this module and nothing in it is decorative. The two bookend blank pages are what make the session honest: without them you finish with a feeling, and with them you finish with two documents and the difference between them. The predictable collapse is block three eating the rest, so it gets a timer of its own.

Why the shape is that shape

Every block is one of the effects from this module. Retrieval before review, generation before instruction, small hints, fading, interleaving, spacing baked into the closing line. There is nothing decorative in it.

The two bookend blank pages are what make the session honest. Without them you finish with a feeling. With them you finish with two documents and a difference between them.

Adapting it

- **25 minutes.** Blank page 3, gaps 3, work 12, retrieval 5, close 2. Drop the break.
- **Two hours.** Run it twice with different topics, not once at double length. Attention does not survive the second half of a long block, and two topics buys you interleaving between them.
- **Problem-set evening.** Replace the retrieval block with more hint-laddered problems, but keep both blank pages — for problem work, write the *method* from memory, not the content.
- **The night before.** Do not do this. Do past papers under time pressure, which is Module 8.

The log is not optional

Three lines at the end of each session, in one file. Date, topic, what you can now do unaided, what you cannot.

It earns its keep four ways. It is your spacing schedule, since *what I cannot do* is next session's opening. It is your honest record of progress, which protects against the fluency trap. It is the raw material for revision, because by exam time you have a list of every genuine difficulty you met. And it is a process record, which Module 7 will explain matters more than any argument you can make after an accusation.

A plain text file is enough. `2026-09-08, enzyme kinetics: can derive Michaelis-Menten and read a Lineweaver-Burk plot unaided; still cannot explain why the double-reciprocal plot distorts error; review Wednesday.`

The failure mode to watch for

The session collapses in one predictable way: block three expands and eats the rest. Working through gaps with an assistant is engaging, the answers keep coming, and forty minutes disappear. You end with a good understanding of one gap and no retrieval practice at all.

Set a timer for block three specifically. When it goes, stop, even mid-problem — write down where you were and come back next session. An unfinished problem is a better start to the next session than a finished one anyway, because you begin with retrieval already in progress.

What this costs

Nothing. A timer, a notebook or plain file, whatever free tier you have, and fifty minutes. The whole structure works with a local model on an offline laptop, and blocks one, two and five work with no computer at all.

The expensive part is not the tools. It is being willing to spend the first five minutes producing something thin and disappointing, every single time, when

a fluent explanation is one keystroke away. That is the whole discipline, and everybody finds it uncomfortable.

BEFORE YOU MOVE ON

What do the two blank-page blocks at the start and end of the session provide that the rest of the structure does not?

A. A record of the session for disclosure purposes, showing that the work in between was the student's own.

B. Spaced repetition within the session, since writing the same content twice at an interval strengthens it more than writing it once.

C. A measurable before-and-after that replaces the feeling of having understood with two comparable documents.

D. Time away from the assistant, which prevents the over-reliance that makes the rest of the session less effective.

---

27 · 8 min

Study groups, with and without the machine

THE ONE THING TO KEEP

Answer alone, argue with someone who disagreed, then check the source — a group that consults the assistant first has an answer and stops producing the reasoning that the exercise exists to train.

Why groups outperform solitary study, when they do

A study group works for one reason: somebody has to explain, and somebody has to be unconvinced. Everything else about them — the sharing of notes, the moral support, the biscuits — is secondary to that exchange.

The evidence is strongest for a specific structure. In **peer instruction**, developed by Eric Mazur for physics teaching, students answer a conceptual ques-

tion alone, then discuss it with a neighbour who chose differently, then answer again. The proportion answering correctly rises substantially after discussion, including in groups where nobody initially had it right — which tells you the gain comes from articulating and challenging reasoning, not from copying the right answer off the strongest student.

That structure survives the arrival of AI intact, and is worth protecting, because the default group behaviour now is for everyone to consult the same assistant and converge silently.

Running peer instruction with three friends and a free model

1. Generate ten conceptual questions from the material — the kind with a common wrong answer, not recall. Ask for *questions where the most popular wrong answer reveals a specific misconception*.
2. Everyone answers question one alone, in writing, with a confidence rating. No discussion.
3. Reveal answers to each other. If you disagree, argue. If you all agree, argue anyway: one of you explains, another must try to break it.
4. Answer again.
5. Only now check against the source or the model.

The order matters. A group that checks first has an answer and stops thinking. A group that argues first produces the reasoning, which is the thing being trained.

The role worth rotating

Appoint one person per session as the sceptic. Their job is to be unconvinced and to ask *why* until either the explanation reaches bedrock or it breaks. Rotate it, because it is the position that learns the most and it is uncomfortable.

This is also the role a model can play when you are alone, using the sceptic prompt from Module 2 — but with a group you get something the model cannot give: a person who is genuinely confused in a way you did not anticipate, and whose confusion is real rather than performed.

Dividing work: the thing that quietly ruins groups

The most common group arrangement is to split the topics — you do chapters one and two, I do three and four, we share notes. This is efficient and almost worthless for learning. Everybody ends with genuine understanding of half the material and someone else's summary of the other half, which is exactly the two-tier outcome the fluency lesson warns about.

If you split anything, split the *teaching* rather than the *learning*: everybody reads everything, and one person is responsible for leading the discussion on each part. Preparing to teach a topic is one of the strongest study conditions known, and it only works if the rest of the group is prepared enough to ask hard questions.

Where a shared assistant genuinely helps

- **Settling factual disputes fast** so the argument can move to the interesting part. Verify anything that matters.
- **Generating the question set**, which nobody enjoys writing.
- **Playing an adversary** for a presentation rehearsal: paste the argument, ask for the hardest questions a panel would ask, take them in turn.
- **Translating between the group's languages.** In a multilingual group, having each person explain in the language they think in and translating for the others is faster and more accurate than everybody working in a second language. This is one of the genuinely excellent uses.

The group agreement

Write it in one message, at the start, before the first deadline:

> *For this project: AI for question generation, critique and translation. No AI-generated prose in the submitted document. Everyone opens every source they cite. We each write our own sections. If anyone uses it differently, they say so before submission.*

Group work is where academic-integrity cases most often begin, and they usually begin because nobody discussed it. You are jointly submitting a docu-

ment, and *I did not know he generated that section* is a poor position to be in, both procedurally and with your friend. Fifteen minutes at the start prevents it.

When there is no group

Distance students, part-time students, people with jobs and families — the group may not exist. The substitutes, in descending order of value: a single study partner on a video call once a week; an asynchronous partner who reads and challenges your explanations in writing; and, last, the model playing sceptic. The last is the weakest, because it will never be genuinely confused, and being confronted with real confusion is a large part of what makes the exchange work.

BEFORE YOU MOVE ON

In peer instruction, groups where nobody initially chose the correct answer still improve after discussion. What does that finding rule out?

A. That the improvement comes from adopting the right answer from a better-informed peer.

B. That confidence ratings are needed for the technique to work, since a group with no correct answer has no confident member to follow.

C. That prior knowledge matters, since a group with no correct answer clearly had none and improved regardless.

D. That the questions were well designed, because a good conceptual question should produce at least one correct answer in any group of four students.

28 · 8 min

Proving it to yourself

THE ONE THING TO KEEP

Predict your score before you check it, and record specific failures rather than grades — the average gap between what you predict and what you get is the most useful number you can have about yourself.

The only test that counts

Every technique in this module points at one question, and it is worth stating plainly, because it is the thing the tools make easy to avoid:

> *With everything closed, can I do it?*

Not recognise it. Not follow it. Not feel that I could. Do it — produce the derivation, write the paragraph, name the mechanism, solve the problem — on blank paper, with no window open and nobody to ask.

Everything else in study is instrumental to that. And it is measurable, cheaply, at any moment.

The unaided check

The procedure takes ten minutes and belongs at the end of every topic, not the end of the term.

1. Close everything. Phone in another room, not face down on the desk.
2. Set a timer for the length the real task would take.
3. Do the thing. A past-paper question, a paragraph of the essay, a derivation, an explanation.
4. Only then open the source and mark it honestly against the mark scheme if there is one.

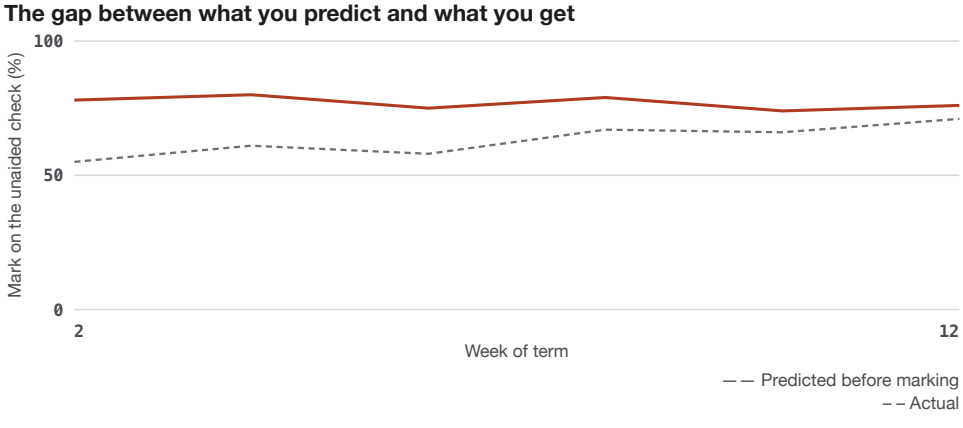
5. Write down the specific failure, not a grade. Not *weak on kinetics* but *could not remember whether the inhibitor changes Km or Vmax and guessed.*

Step five is where the value is. A grade tells you how you did. A specific failure tells you what to do on Wednesday.

The calibration column

Before you check, predict your score. Write the number down.

Over a term this produces the most useful single statistic you can have about yourself: the average gap between predicted and actual. If you consistently over-predict by 15 per cent, you now know that your sense of preparedness runs 15 per cent hot, and you can correct for it — which is worth more than any individual mark, because it tells you when to stop revising a topic and when you were wrong to.



One student's calibration column across a term: the predicted mark written down before marking, against the mark. The average gap, here about fourteen points narrowing to five, is the most useful single number anybody can hold about themselves, because it says how much to discount your own sense of being ready. Students leaning hardest on an assistant tend to over-predict most, which is the fluency trap showing up as arithmetic.

Students using AI assistants heavily tend to over-predict more, for the reason in the fluency lesson. Measuring your own gap is how you find out whether that describes you, rather than assuming it does or does not.

Building a portfolio of things you can do

Keep a single file, appended to after each unaided check. One line each: the date, the task, and whether you completed it unaided.

> *12 Oct — derived the quadratic formula by completing the square, unaided, 6 min.*

> *14 Oct — could not reproduce the proof that root 2 is irrational without a hint. Retry 17 Oct.*

> *17 Oct — reproduced it unaided.*

By the end of a module this file is a list of demonstrated capabilities with dates. It is the answer to *what have I actually learned*, and it is immune to the feeling of competence, because it records outcomes rather than impressions.

It is also, if you ever need one, an extraordinarily good piece of evidence in an academic-integrity conversation — a dated record of you building capability under your own steam. Module 7 returns to this.

The failure this catches early

The pattern to fear is: an excellent term of assisted work, high coursework marks, a comfortable sense of understanding, and then a closed exam where none of it is available. By the time that gap appears in an exam hall it is far too late to close.

The unaided check finds it in week three, when it costs an afternoon to fix. It is a smoke alarm, and like a smoke alarm its value is entirely in going off before the thing you were worried about.

Raising the difficulty deliberately

Once a check is comfortable, make it harder in the direction the real test will:

- **No notes** if you had notes.
- **Timed** if it was untimed.

- **Handwritten** if the exam is handwritten. Writing by hand is slower and more effortful than typing, and if you have not practised it, you will run out of time on material you knew.
- **Cold** — a topic from three weeks ago, unannounced, rather than the one you just revised.
- **Explained aloud** to somebody, if any part of your assessment involves speaking.

Each of these narrows the gap between the conditions you practise in and the conditions you will be assessed in, which is the whole subject of Module 8.

The honest closing note

This is the least enjoyable lesson in the module to act on. Nothing about the unaided check is pleasant: you sit down knowing you might not be able to do it, and quite often you cannot. The assisted alternative is right there, and it feels better, and it produces a document.

But the question at the top of this lesson is the one the world will eventually ask, in an exam hall or a job or the first time something goes wrong and nobody is available. The students who came through this era able to do things are the ones who kept asking it of themselves, early, while it was still cheap to answer badly.

BEFORE YOU MOVE ON

Why is writing down a predicted score before marking an unaided check more useful than the mark itself?

A. Because it commits you to a target, and committed targets improve subsequent performance through goal-setting effects.

B. Because it reduces the disappointment of a low mark by managing expectations in advance.

C. Because predicting first prevents the mark scheme from anchoring your judgement of your own answer, which would inflate your self-assessment.

D. Because the gap between prediction and result measures how far your sense of preparedness can be trusted.

CASE STUDY · MODULE 3

The evening they stopped finishing things

Three students sharing a room at a coaching centre in Kota, eleven weeks from an entrance examination, changing how they studied after a mock that would not move.

Prerna Yadav, Imtiaz Khan and Rohit Meena had an arrangement that looked sensible and had been in place since June. They split the chapters. Each made notes on their third and shared them. Every Sunday the centre ran a mock out of three hundred, and every Monday a board in the corridor listed problems attempted by each student in the previous week.

Their attempted counts were around a hundred and forty each. Their mock scores across four Sundays were a hundred and sixty-two, a hundred and sixty-eight, a hundred and fifty-nine and a hundred and sixty-six. Nothing was moving, and all three of them were working long hours.

Prerna's proposal, after the fourth Sunday, had three parts. Nobody asks the assistant for a solution: hints only, smallest first, and only after six or seven minutes with no forward progress, because struggle with no resolution is not a learning condition, it is an unpleasant afternoon. Conceptual questions get answered alone, in writing, with a confidence rating, before anybody speaks — then whoever disagreed argues it out, then everyone answers again, and only then does anyone check the source. And every night ends with ten minutes on blank paper, alone, reproducing something from earlier in the week.

The costs were immediate and they were not small.

Attempted problems fell from about a hundred and forty a week each to sixty-two. The board in the corridor is read by the staff and by everybody's parents on visiting weekends, and it does not have a column for what any of the sixty-two taught anybody. Imtiaz's father asked him on the phone why he had stopped working.

The mock score fell before it rose. The Sunday after they started, Prerna scored a hundred and fifty-four, which was her worst mark since April. She had expected the dip in the abstract and found it much harder to sit with in

fact, because there were eleven weeks left and no guarantee that the curve turned.

And Rohit had a constraint the other two did not. His scholarship renewal was decided on his mock average across the next five Sundays, and the deadline was the deadline. A method that costs marks now and returns them in six weeks was, for him, a method that cost him the year. He stayed in the arrangement for three weeks, watched his average slide, and went back to reading worked solutions. He told them he was doing it and why. Prerna's account afterwards is that he was right, and that the dip is real and his deadline punished it, and that people who pretend otherwise are not being honest about what the evidence says.

The two who stayed also found that one part of the proposal did not work as written. The peer instruction round only does anything when somebody disagrees, and in a week where all three had answered identically the argument was theatre. They added a rule: if everyone agrees, one person has to try to break the explanation, out loud, properly.

STOP HERE

Before turning the page: what would you have done, and what would it have cost if you were wrong? Write it down. The point of the next section is to compare.

WHAT HAPPENED

Prerna and Imtiaz kept it for the remaining eleven weeks. Her Sunday scores from the change onwards ran a hundred and fifty-four, a hundred and sixty-one, a hundred and fifty-nine, a hundred and seventy-seven, a hundred and ninety-one, a hundred and eighty-eight. The attempted count never recovered and finished around eighty.

The number that changed most was not the score. Prerna kept a second column beside each mock: the mark she predicted before she looked. In the first fortnight she was over by an average of twenty-two. By week nine she was over by six. She says the useful part was not the accuracy but knowing which topics

she was wrong about being ready for, because those were the ones she had never checked and would not have checked.

The ten minutes of blank paper at the end of each night was the part they nearly dropped, three separate times. It is the least enjoyable thing in the arrangement: you sit down knowing you may not be able to do it, and often you cannot, and a fluent explanation is one keystroke away.

Rohit renewed his scholarship. His final entrance score was lower than either of theirs. He does not believe the arrangement is the reason, and neither do they, and with three people and one attempt nobody is in a position to say.

Their scores fell for three weeks before rising. Name the mechanism, and explain why a method that lowers performance during practice can raise it afterwards.

Desirable difficulties. The conditions that make performance during study feel good and look good are frequently the ones that produce the least durable learning, and the reverse holds too. Retrieval from a blank page, spacing sessions so that a little forgetting happens in between, and interleaving problem types so that nobody tells you which method to use are all slower and more error-strewn in the session and better a week later. Their attempted count fell because they were no longer being carried through problems; their early scores fell because they had removed the scaffolding before the capability had replaced it. The rise is not the arrangement finally working. It is the arrangement having been working the whole time on something the Sunday mock did not measure until it did.

Was Rohit's decision wrong? Say what it cost him and what it bought him.

It was not wrong, and treating it as a failure of nerve misdescribes his situation. He faced a five-week window in which the dip is measured and the recovery is not, with a scholarship attached. Taking solutions bought him the marks in that window and it bought them at a known price: he spent those five weeks reading correct derivations, which teaches the shape of a solution and very little about constructing one, so the capability he might have been building was deferred. The course's own advice covers this case explicitly. A solution taken under pressure is not a failure unless you never come back to it, and the repair is to mark the problem and redo it cold a few days later. The mistake available to him was not taking the solutions; it was not returning.

Attempted problems fell from a hundred and forty to sixty-two and the score rose. What was the hundred and forty measuring, and why did the centre put it on a board?

It was measuring throughput, which is visible, countable and weakly related to what the examination tests. A board can carry it because it requires no judgement to record. What it cannot distinguish is a problem solved and a problem watched being solved, and once an assistant is in the room the second is much faster than the first, so the metric rewards exactly the behaviour that costs marks later. The honest measure of a study session is what you can produce on blank paper afterwards, which is why the arrangement ended each night with ten minutes of it. That measure is slower to collect and nobody can put it on a board, which is most of why boards carry the other one.

MODULE 3 · TEST

Check what stuck

6 questions, one right answer each. This one is for you, not for a record: answers and the reason behind each are at the back. If two questions from the same module go wrong, re-read that module before the exam rather than after.

1 Roediger and Karpicke's re-readers won the test five minutes later and lost the one a week later. What does that show about retrieval practice?

A. Rereading works, provided it is done at least four times
B. Recall helps only when the material is already well learned
C. Retrieval leaves you better prepared and feeling worse
D. Testing measures what you know without changing what you know

2 You have an answer that does not match and want help locating the error. Which clause preserves the learning?

A. Tell me which line is wrong, not what it should be
B. Give me the full solution, then explain each step in turn
C. Tell me how confident you are before you give the hint
D. Summarise the method before you look at my working at all

3 Interleaved practice scores worse during the session and better a week later. What is the mechanism?

A. Mixed practice covers more of the syllabus in the same hours
B. Switching topics keeps attention higher across a long session
C. It spreads the work, so each topic is revisited more often
D. In a block the heading tells you which method to use

4 What does the expertise reversal effect say about worked examples?

A. Examples help everybody, so read as many as you can find
B. They help beginners and hinder those who have the method
C. They help only in mathematics, where the steps are formal
D. They work best when read after attempting the problem cold

5 You write down the mark you expect before checking, every time, for a term. What is that number for?

A. It converts practice scores into a predicted final grade

B. It tells your teacher which topics the class found hardest

C. It says how much to discount your own sense of readiness

D. It replaces the mark scheme when none has been published

6 In peer instruction, the group checks the source only after arguing. Why does the order matter?

A. A group that checks first stops reasoning

B. Checking first wastes the group's shared message allowance

C. The source is more convincing once everybody already agrees

D. Sources are unreliable until several people have read them

MODULE 4

Subject by subject

The same tool behaves differently in maths, code, the lab sciences, the essay subjects and language learning. Each has a characteristic failure, and each has one check that catches it.

> BY THE END OF THIS MODULE YOU CAN
>
> Name the characteristic failure mode of an AI assistant in a given subject — arithmetic in maths, plausible invalid steps in proof, code that runs and is wrong, invented quotations in history, fossilised errors in language learning — and apply the specific check that catches it, using a free tool where a deterministic one is required

29 · 9 min

Maths, and the arithmetic problem

> THE ONE THING TO KEEP
>
> Digits are tokenised without place value, so arithmetic degrades with length while looking perfectly normal — take the method from the model and the number from SymPy, Maxima or a code interpreter.

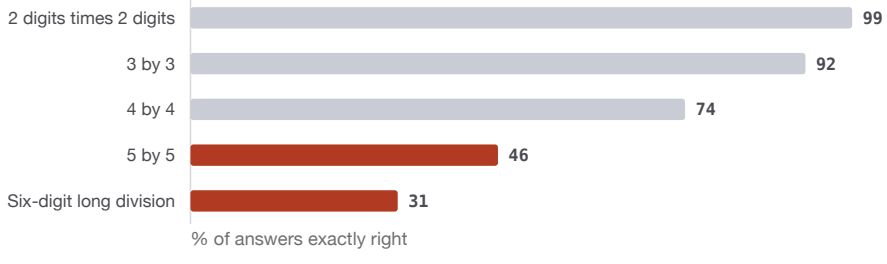
Why a system that discusses topology gets 47 times 63 wrong

It is genuinely strange the first time you see it. A model that can explain the intuition behind a manifold will misplace a digit in a three-digit multiplication. The reason goes back to tokens.

Numbers get split into tokens in ways that do not respect place value. Depending on the tokenizer, 4763 might be `47` and `63`, or `4`, `76`, `3`. The model never sees a column of digits to carry between. It has learned enormous numbers of arithmetic examples and produces the most plausible continuation, which for small familiar numbers is almost always right and for larger or unusual ones is a good guess.

So performance degrades with digit count in a way no calculator's ever would. Two-digit multiplication is usually fine. Four-by-four is unreliable. Long division, arithmetic with many decimal places, and anything requiring exact rational arithmetic are all risky, and the failures look completely normal — a confident wrong number in the middle of correct working.

Accuracy falls with digit count, and the tone does not



The shape rather than a benchmark. Numbers are split into tokens that do not respect place value, so there is no column of digits to carry between and the model produces the most plausible continuation, which for small familiar numbers is nearly always right. The failures look exactly like the successes: a confident wrong number in the middle of correct working. Take the method from the model and the number from SymPy, Maxima or a code interpreter.

The rule

Never accept a number a model computed if the number matters. The concept, the method, the setup, the interpretation: all fine. The arithmetic: check it.

This is not a limitation to work around with clever prompting. It is what the machine is.

Free tools that do the arithmetic properly

- **A code interpreter**, if your assistant has one. This is the real fix: the model writes Python and the Python runs. The answer then comes from an

actual computation. Check that the code is right, because the code can be wrong; but if the code is right, the number is right.

- **SymPy** — free Python library, symbolic algebra, exact arithmetic, calculus, equation solving. Runs in Google Colab with no install.
- **Maxima** — free, mature computer algebra system, decades of use.
- **wxMaxima** for a graphical front end.
- **GeoGebra** and **Desmos** — free, browser-based, excellent for graphing, geometry and checking that a function does what you think.
- **GNU Octave** — free and runs most MATLAB code, which matters because job ads name MATLAB.
- **SageMath** — free, covers a lot of Mathematica's territory.

The pattern that works: ask the model *which* method, ask it for the SymPy or Maxima code, run it, compare with your own answer.

```
from sympy import *
x = symbols('x')
# check an integral you did by hand
integrate(x*exp(x), x)          # x*exp(x) - exp(x)
# and confirm by differentiating back
diff(x*exp(x) - exp(x), x)      # x*exp(x)
```

That is thirty seconds and it is a real check, because SymPy is not guessing.

Where models are genuinely strong in maths

Do not overcorrect. There is a lot they do well:

- **Explaining what a definition means** and why it is stated that way.
- **Giving the intuition** behind a theorem before the formal statement.
- **Suggesting an approach** when you are stuck: *this looks like a substitution problem, try u equals the inside of the bracket.*
- **Generating practice problems** of a specified type, and hint ladders for them.

- **Spotting the step where your working goes wrong**, which they do well and which is the hint-ladder move from Module 3.
- **Translating between notations**, when your lecturer uses one and the textbook another.

That is most of what a maths tutor does apart from the arithmetic.

Word problems and the setup

The hardest part of an applied maths question is usually turning the words into equations, and this is a genuinely good use. Ask for the setup only:

> *Turn this into equations. Define every variable with units. Do not solve it.*

Then solve it yourself. You have removed the part you find hard for the wrong reasons — parsing dense prose — and kept the part the exam is testing.

Checking without any tool

When you have no computer, three checks cost nothing:

Estimate first. Round everything hard and get an order of magnitude before you look at the answer. 47 times 63 is about 50 times 60, so about 3000. An answer of 296 or 29,610 is caught instantly.

Units. Carry them through every line. If they do not come out as they should, the error is above that line.

Substitute back. Solved for x? Put it in the original equation. This catches most algebra errors and takes fifteen seconds.

These are the same checks a good teacher has always taught, and they matter more now, not less, because the source of the wrong number has changed from tiredness to something that never gets tired and is wrong anyway.

BEFORE YOU MOVE ON

A student uses a model to solve a physics problem. The setup and reasoning are correct, but the final figure is 2.94 kJ where the true answer is 29.4 kJ. Which check would have caught this fastest?

A. Asking the model to double-check its arithmetic, since re-examination usually surfaces slips.

B. A rough order-of-magnitude estimate made before looking at the answer.

C. Carrying units through every line, since a factor-of-ten error changes the dimensions of the result.

D. Asking the same question again in a fresh conversation and comparing the two answers.

30 · 8 min

Proofs, and the step that looks fine

THE ONE THING TO KEEP

Generated proofs fail by making one unjustified move inside otherwise correct working, so read upward from the conclusion asking what licenses each step, and check the seven standard illegal moves.

A proof is a chain with no weak links permitted

Most writing tolerates a slightly loose sentence. A proof does not. One unjustified step and the whole thing is worth nothing, however correct the conclusion happens to be. This makes proof the subject where generated text is most dangerous, because a model produces *proof-shaped* text extremely well.

The characteristic failure is not gibberish. It is a proof that is correct for six lines, makes one move that is not justified, and continues correctly to the right answer. Everything reads normally. Markers find these instantly, because an

unjustified step is exactly what they are trained to look for, and it is the thing being examined.

The moves that go wrong

A short list covers most of them:

- **Dividing by something that might be zero.** The classic, and it appears constantly.
- **Assuming what is being proved**, usually several lines in and rephrased so it does not look like it.
- **Swapping quantifiers.** *For all x there exists y* is not *there exists y for all x*, and the difference is the content of half of analysis.
- **Using a theorem outside its hypotheses.** Applying a result about continuous functions to one that is not, or a result about finite sets to an infinite one.
- **Taking a limit inside something** — a sum, an integral, an expectation — where that requires justification.
- **Induction with a broken step.** The base case is checked, the inductive step quietly assumes the statement for all smaller n when the induction was set up for n minus one only.
- **Special case treated as general.** Verifying three examples and writing *hence in general.*

Read any generated proof looking specifically for these seven, in this order. It takes two minutes and catches most of what goes wrong.

Read it backwards

A technique from mathematicians that works especially well on generated text. Start at the conclusion and work upward, asking of each line: *what exactly justifies this from the line above?* Forward reading rides the momentum of a plausible narrative. Backward reading breaks it, because you are forced to supply the justification yourself rather than accepting that one was implied.

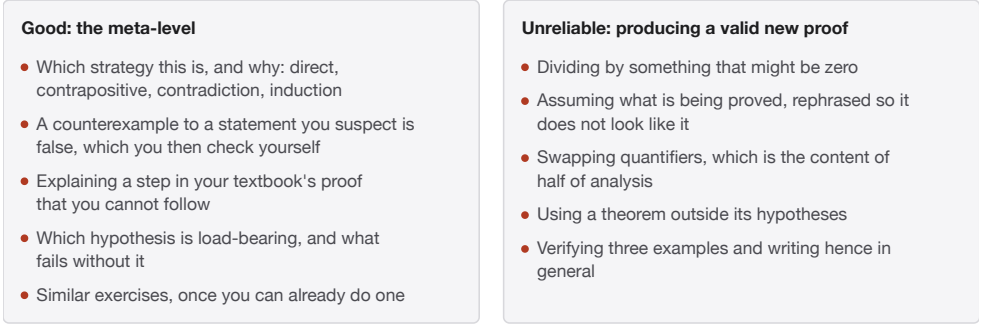
If you cannot name the rule that licenses a step, that step is not proved, whatever it says.

Where a model is actually useful in proof work

- **Proof strategy.** *Is this more likely to be direct, contrapositive, contradiction, or induction, and why?* This is genuinely good advice most of the time, and choosing the strategy is where beginners get stuck.
- **Finding a counterexample** to a statement you suspect is false. Then verify the counterexample yourself, which is easy — checking a counterexample is much cheaper than finding one.
- **Explaining a step in a proof from your textbook** that you cannot follow. Grounded on a real proof, this is safe and very useful.
- **Asking what the theorem is really doing** — which hypothesis is load-bearing, what fails without it. Excellent for building intuition.
- **Generating similar exercises** once you can do one.

Notice the pattern: the model is good at the meta-level — strategy, intuition, explanation of an existing proof — and unreliable at producing a valid new proof.

What a model is for in proof work, and what it is not



The characteristic failure is not gibberish. It is six correct lines, one unjustified move, and the right conclusion. Read upward from the conclusion asking what licenses each step: forward reading rides the momentum of a plausible narrative, and backward reading breaks it because you have to supply the justification yourself. If you cannot name the rule, that step is not proved.

Proof assistants, and the free path

If you want a machine that genuinely cannot accept an invalid step, that exists and is free: **Lean** (with its mathlib library), **Coq**, and **Isabelle** are proof assistants that check every inference mechanically. They are hard to learn and out of scope for most undergraduate courses, but the existence of the distinction is worth knowing: a language model produces text that resembles a proof, and a proof assistant verifies a proof. Those are different categories of thing, and confusing them is the error underneath every story about an AI-generated proof that turned out to be wrong.

Lean has an active community and a free online textbook, *Mathematics in Lean*. If you are heading towards pure maths, an afternoon with it will permanently change how you read a proof, because it forces you to notice every step you had been waving at.

Writing proofs for assessment

The honest position: reading generated proofs teaches you the shape of proof writing and does not teach you to construct one, for the same reason reading worked examples does not teach you to solve problems. If your assessment includes proofs, the unaided check from Module 3 is not optional. Write one from blank paper, weekly, timed.

And when you are stuck, use the hint ladder rather than the solution. *What kind of proof is this likely to be?* is rung two. *What is the first line?* is rung three. The rung that gives you the whole argument is the one that teaches you nothing at all.

BEFORE YOU MOVE ON

Why does reading a generated proof backwards from the conclusion catch errors that forward reading misses?

A. Because errors are more likely near the end of generated text, where the model has more context to be distracted by.

B. Because a proof read backwards is logically equivalent to its contrapositive, which is easier to verify.

C. Because forward reading rides a plausible narrative; upward reading makes you supply each justification.

D. Because the conclusion is the only line whose correctness you can check independently, so verification must start there.

31 · 9 min

Code that runs and is wrong

THE ONE THING TO KEEP

Generated code that runs is not code that is correct — read every line, test the boundary case, refuse rewrites when debugging, and write the tests yourself from the specification.

The most dangerous output in this course

Generated code has a property nothing else in this course has: it will happily run, produce plausible output, and be wrong, and you will only find out later. A wrong essay sentence gets marked. A wrong line of code gets a result you then build on.

Students new to programming are the most exposed, because the failure requires you to read the code to detect, and reading code is precisely the skill not yet acquired.

The characteristic failures

Off-by-one and boundary errors. Loops that miss the last element, ranges that include an endpoint they should not. The code runs; the answer is slightly wrong.

Invented APIs. A function, method or parameter that does not exist in the library, with a completely plausible name. This is fabrication in a different costume, and it is common enough to have its own security concern: attackers register package names that models frequently hallucinate. If a package name comes from a model, check it exists on the real index before installing it.

Version drift. Correct code for an older version of a library. Deprecated arguments, renamed functions, changed defaults. The model's training has no clean sense of which version you have installed.

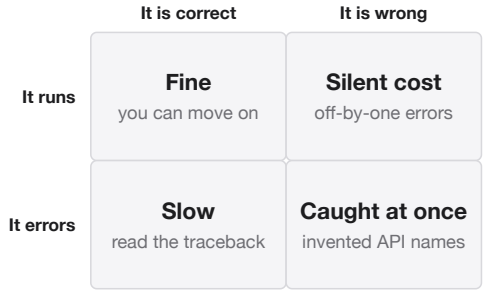
Silent type problems. Integer division where you wanted float, string concatenation where you wanted addition, a pandas operation returning a copy when you expected a view.

Working code, wrong problem. The most expensive one. It solves what the model understood you to mean, which is not what you meant, and it works perfectly.

The rule

Read every line before you run it, and be able to say what each line does. If you cannot, you have not received help; you have received a liability. Run it, then test it on a case where you already know the answer — including an edge case: empty input, one element, a negative number, a duplicate.

The cell that costs you is the quiet one



Code that runs is not code that is correct. Errors announce themselves and cost a minute; the off-by-one, the integer division and the program that solves what the model understood you to mean all run cleanly and get built on. Read every line and be able to say what each one does, then test a case where you already know the answer, including empty input and a single element.

That last habit is worth more than any other in this lesson. Generated code is usually tested by the person who received it against exactly the case they had in mind. Testing the boundary takes ten seconds and catches most of the off-by-one class.

Debugging, which is the genuinely excellent use

This is where these tools are transformative and where using them well is unambiguously good.

> *Here is the error, the full traceback, and the relevant function. Do not rewrite my code. Tell me what the error means, and what would cause it here. I will fix it.*

Do not rewrite my code is the essential clause. Without it you get a rewritten file and no idea what was wrong. With it you get an explanation, and the ability to recognise that error unaided next time, which is the actual skill — professional programmers spend far more time reading errors than writing new code.

Also excellent: pasting a piece of unfamiliar code and asking what it does, line by line. This is reading practice with an infinitely patient annotator, and it is one of the fastest ways to get comfortable in a new codebase or language.

Version control is your process record

git is free, and for a student it does two jobs at once. It is how you avoid losing work, and it is a dated, granular record of your work developing — which is exactly the evidence Module 7 says you want to have before anyone asks for it.

The minimum that helps:

```
git init
git add -A && git commit -m "first attempt at the parser, fail-
ing on nested cases"
```

Commit at the end of every session, including the bad ones. A history of forty commits showing an idea developing, with failures in it, is worth more than any argument you could construct afterwards. GitHub and GitLab both host free private repositories.

What to keep doing yourself

For a computing student, this list is not optional:

- Writing a function from a specification, unaided, regularly.
- Reading a stack trace and locating the fault before asking anything.
- Reasoning about complexity — why this is quadratic and that is linear.
- Writing the tests. If the model writes both the code and the tests, the tests check the code's assumptions rather than the specification's.

That last point is subtle and important. Tests written by whatever produced the code will pass by construction. Write the tests first, yourself, from the spec, then let the model help with the implementation. That order preserves the check.

Free tools, named

Python, VS Code or **VSCodium** (the same editor without the telemetry), git, and **Google Colab** for a free notebook with GPU access. For learning to read errors, no tool beats `print` statements and a debugger you have actually

opened. Copilot-style completion in the editor is convenient and, for a learner, actively risky: it produces code before you have formed the intention, which removes the design step from the task. Turn it off while you are learning the language, and turn it on later when you can tell instantly whether what it suggested is right.

BEFORE YOU MOVE ON

Why is it a problem to let a model write both the implementation and the tests for a piece of coursework?

A. Because generated tests tend to be shallow and only cover the obvious path through the function.

B. Because tests written after the implementation are always weaker than tests written first, regardless of who writes them.

C. Because most marking schemes award credit for tests separately, so generated ones cost marks directly.

D. Because the tests encode the same misunderstanding as the code, so they pass by construction.

---

32 · 8 min

The lab sciences, and the dimension check

THE ONE THING TO KEEP

Carry units through every line as algebra and the dimensions will catch most wrong formulae instantly — then apply significant figures and uncertainty yourself, because those are the parts a lab report is actually marked on.

Where physics and chemistry problems break

In a quantitative science, the model's weaknesses land on the parts where marks are actually allocated: the number, the unit, the significant figures, and

the assumption that made the calculation legitimate.

The good news is that these subjects come with the strongest error-detection method in this entire course, and it is free, instant and does not need a computer.

Dimensional analysis, used properly

Carry the units through every line as though they were algebra. Cancel them. Whatever remains must be the unit of the answer.

This catches a startling proportion of errors, including most of the ones a generated solution will contain, because a wrong formula usually has wrong dimensions.

> *Kinetic energy: half m v squared. kg times (m/s) squared is kg m² s⁻², which is a joule. Correct.*

> *If a solution produced kg m s⁻², that is a newton, and the error is upstream — a v that should have been v squared, most likely.*

Dimensional analysis also lets you reconstruct formulae you have forgotten. If you know a period depends on length and gravity, the only combination of metres and m s⁻² giving seconds is the square root of length over g. You have just recovered the pendulum formula without remembering it, up to a constant.

Ask the model to do this deliberately:

> *Before giving me a number, show the dimensions of every quantity and confirm the units of the result.*

Significant figures, which are marked and usually wrong

Models routinely give ten decimal places or round to two when the data supported four. Both lose marks. The rules are yours to apply:

- Multiplication and division: the result carries the same number of significant figures as the least precise input.

- Addition and subtraction: the same number of decimal places as the least precise input.
- Constants and exact counts do not limit precision.
- Round once, at the end, not at each step.

This is a place where a marker can see, in one glance, whether the person understood their measurement or copied a number. It is cheap to get right and it is regularly the difference between full and partial credit.

Uncertainty, which is the actual physics

A measurement without an uncertainty is not a result. Models handle uncertainty propagation adequately in simple cases and badly in correlated ones, and they will often omit it entirely unless asked.

For a lab report, do it yourself. The rules for independent uncertainties are short: add absolute uncertainties for sums and differences, add relative uncertainties in quadrature for products and quotients. Ask the model to explain *why* those rules take that form — the derivation from a first-order expansion is genuinely worth understanding — and then apply them by hand or in a spreadsheet.

Free tools: LibreOffice Calc or Google Sheets for propagation, **Python with NumPy** for anything larger, and the `uncertainties` package if you want it done automatically.

Chemistry specifics

- **Balancing equations.** Check every element yourself; a plausible-looking unbalanced equation is common.
- **Oxidation states and mechanisms.** Reasonable on standard textbook cases, unreliable on anything unusual, and confidently wrong on stereochemistry.
- **Structures and diagrams.** Text descriptions of molecular geometry are frequently muddled. Draw it. **Avogadro** and **Jmol** are free molecular editors; **ChemDraw** is what industry names, and **ChemSketch** has a free academic version.

- **Safety information.** Never take a hazard, incompatibility or first-aid answer from a model. Use the actual safety data sheet, which is free and legally required to exist for every reagent in your lab.

Biology and the currency of specifics

Biology is unusually exposed to the specificity problem: it runs on named genes, named pathways, named species, named studies and constantly-revised guidelines. A model is strong on mechanism at textbook level and unreliable on any of those particulars, especially anything revised since its training.

And the standing rule for anything clinical: **do not use it for advice about a real person, including yourself.** Guidelines change, they differ by country, and the model has no way to know your case. Use it to understand a mechanism you are being taught. Ask a clinician about a patient.

What the model is genuinely good at here

Explaining why a formula has the form it has. Building intuition for a concept before the maths. Generating practice problems of a named type. Checking your working for the *first* wrong line. Explaining a paper's methods section. Suggesting what could have caused an anomalous result in your data, as a list of hypotheses for you to test — which is a real and underused move for a discussion section, provided you evaluate the hypotheses rather than reporting them.

BEFORE YOU MOVE ON

A student checks a generated solution and finds the units come out as kg m s^{-2} where the answer should be in joules. What have they learned?

A. That an error exists upstream — a quantity used to the wrong power or a wrong formula.

B. That significant figures were applied at the wrong stage of the calculation.

C. That the final numerical value needs converting into the correct unit before submission.

D. That the model used a different but equivalent formulation, since force times distance and energy are dimensionally related quantities in some conventions.

33 · 8 min

History, and the quotation that was never said

THE ONE THING TO KEEP

History runs on exactly the specifics a model is weakest at, so trace every quotation to a primary source — a phrase search in Google Books restricted by date will expose most invented attributions in two minutes.

The subject where fabrication is invisible

In maths a wrong answer is wrong. In history a fabricated detail looks exactly like a real one, because the whole subject is made of specifics: dates, names, quotations, archive references, the position a particular historian took in a particular book.

Those specifics are the weakly-represented category from Module 1. Which means history is the subject where the mechanism bites hardest and where the failure is hardest to see without checking.

The four fabrications to expect

Invented quotations. The most common and the most damaging. A plausible line, attributed to a real figure, in their register. Famous historical figures attract these enormously — many quotations widely attributed to Churchill, Gandhi, Lincoln or Einstein were never said by them, and that misattribution is in the training data too, so the model reproduces it confidently.

Composite events. Two similar events blended into one, with a date from one and details from the other. Reads perfectly. Wrong.

Invented historiography. *Historian X argued that…* — where X is real, the position is plausible for their school, and they never wrote it. Markers in history departments spot this immediately, because they know the literature.

Date drift. Confident dates that are one or two years out. Especially bad for anything before 1800, and for events outside western Europe and North America, where the training text is thinner.

The check, which is the discipline of the subject anyway

History already has the answer: go to the source. What is new is that you must apply it to a category of claim that did not previously exist — the claim that arrives already footnoted-looking.

- **Every quotation gets traced to a primary source or a scholarly edition.** If you cannot find where it was said, written or recorded, you do not use it. A quotation you found only in secondary repetition is not evidence.
- **Every historian's position gets traced to the book or article**, with a page number, that you have opened.
- **Every date gets checked** against a reference work.

The free path is good here. **Google Books** for full-text search of published works, which is excellent for finding whether a quotation appears anywhere before a given year. **Internet Archive** and **HathiTrust** for out-of-copyright texts. **JSTOR** has free limited reading; your library has the rest. National

archives are increasingly digitised and free — the UK National Archives, the US National Archives, and many national and state collections.

The bibliographic test for a quotation

A specific and effective move. Search the exact phrase in Google Books, restricted by date. If a quotation attributed to a nineteenth-century figure first appears in print in 1987, it is a twentieth-century invention. This is how the quotation-checking community works and you can do it in two minutes, free.

Where the model earns its place in history work

Substantially, and it is worth being clear about it:

- **Explaining a historiographical debate's shape** — what the positions are, roughly why they differ. Then go and read the actual historians.
- **Context for a primary source you are reading.** Paste the source; ask what a reader at the time would have understood by a term. Verify anything specific.
- **Translating and paraphrasing archaic or foreign-language text**, which is genuinely strong, with the caution that a translation you cannot check is a claim you cannot verify.
- **Arguing against your thesis**, which is the essay move from Module 2 and is worth more in history than almost anywhere else.
- **Making your prose clearer**, at sentence level, on writing you produced.

Source criticism has not changed

The questions you were taught still apply and now apply to a new kind of text. Who produced this? For what audience? What did they gain by saying it? What would they not have known? What is missing?

Run those on a model's output and the answers are unusual: produced by a statistical process, for you, gaining nothing, knowing whatever was in its training text, missing everything specific that was rare. That is not a source. It

is a starting point with no provenance, and treating it as a source is the single mistake to avoid in this subject.

The one that gets people caught

A fabricated citation in a history essay is the most common academic-integrity trigger in humanities departments, and it usually was not intended as dishonesty — the student believed the reference was real. The intention does not change the outcome. **Open every source.** Ten minutes with a library catalogue is the entire defence, and it is also just doing the subject properly.

BEFORE YOU MOVE ON

A student finds a striking quotation attributed to a nineteenth-century reformer. A full-text search shows its earliest appearance in print is a 1974 popular history. What follows?

A. It is usable if cited to the 1974 book, since that is a published secondary source making the attribution.

B. It is probably a later invention and should not be used as evidence of what the reformer said.

C. It is likely genuine but from an unpublished manuscript, since full-text search only covers printed books.

D. The search proves nothing, because digitisation coverage before 1900 is too patchy for absence to be meaningful.

34 · 8 min

Literature, and the text it half remembers

THE ONE THING TO KEEP

Generated quotations from literary works are usually near-misses, and an argument built on words that are not in the book is worth nothing — read and notice first, consult the critical consensus only afterwards.

It has read about the book

A model has encountered enormous quantities of writing *about* literary works — reviews, essays, study guides, forum arguments — and, for older works, often the text itself. What it does not have is your reading of it, and it cannot do the one thing literary study is actually about: attending closely to particular words on a particular page.

This produces a specific failure. Ask about a well-known novel and you get a competent summary of the critical consensus, which is what a study guide would give you, and which is what a marker gives low marks to. Ask for a quotation and you will often get a plausible paraphrase presented as the text.

Misquotation is the characteristic error

Generated quotations from literary works are frequently near-misses: the right sense, the wrong words, or words assembled from different parts of the text. For a subject where the argument is built on the exact wording, this is fatal. A close reading of a sentence that is not in the book is worth nothing, and it is immediately visible to anyone who knows the work.

Every quotation gets checked against the text in front of you, with a page or line reference. Not against a search result. Against your copy, or a reliable scholarly edition. For out-of-copyright works, **Project Gutenberg** and the **Internet Archive** are free and searchable; for verse, be careful about which edition you are using, since line numbering differs.

What close reading is, and why it cannot be delegated

Close reading is noticing that a word is doing something unexpected, and building an argument from it. The noticing is the skill. It requires having the text in front of you, slowly, and being willing to sit with a sentence that will not resolve.

A model can tell you what critics have noticed. That is genuinely useful *afterwards* — it tells you whether your observation is original, obvious, or contested. It is useless *before*, because reading the consensus first means you will notice what you were told to notice, and the resulting essay reads exactly like what it is.

The order that works: read, notice, write your observation, *then* consult. Never the reverse.

Uses that are genuinely good

- **Historical and linguistic context.** What did this word mean in 1798? What would a contemporary reader have understood by this allusion? Strong, and verify the specifics.
- **Prosody and form.** Scanning a line, identifying a form, naming a device. Reliable on standard forms and worth checking against the poem, since it is working from a possibly-misremembered text.
- **Foreign-language and older texts.** Getting a foothold in Chaucer, or in a language you read imperfectly. A ladder to the original, not a replacement for it.
- **Counterargument.** *What is the strongest objection to this reading?* Excellent, and it is where most literature essays are weakest.
- **Structural feedback on your own draft.** Which paragraph does not advance the argument, where the evidence is thin. Ask for diagnosis, not rewriting.
- **Finding secondary literature to look up.** Ask what critical debates surround a work, then find the actual articles yourself. Treat every named article as unverified until you hold it.

Study guides, old and new

SparkNotes and its descendants have existed for decades and departments have always known what an essay written from one looks like. A model produces the same thing more fluently. The tell is identical: an essay that is *about* the novel's themes at a distance, with no sustained attention to any actual passage, and no observation that could only have come from having read it.

The strongest thing you can put in a literature essay remains a small, precise, surprising observation about a few words on the page, defended. No tool can supply that, because it depends on a reading nobody else has done.

For set texts and exams

Many literature exams are closed-book or use a clean text. Memorising quotations is a real requirement, and it is one place where the model is a good drilling partner: paste twenty quotations you need, ask to be tested on them one at a time with the first three words as the prompt, and to be told which you keep getting wrong.

That is retrieval practice on material you have verified yourself, which is the safe and effective arrangement — the verification happened first, and the drilling happens on text you know is right.

BEFORE YOU MOVE ON

Why does consulting a model about a novel's themes before doing your own close reading damage the resulting essay?

A. Because the model's account of the themes is usually wrong, and the misreading propagates into the essay.

B. Because markers can detect the stylistic signature of model-influenced argument even when the student writes every word themselves.

C. Because quotations supplied at that stage will be unverified and are likely to be misremembered.

D. Because you will then notice what you were told to notice, and the essay becomes a version of the consensus.

35 · 9 min

Languages, where it is genuinely excellent

THE ONE THING TO KEEP

The tool is genuinely excellent for input, explanation and drilling, but an agreeable partner that understands your errors without flagging them will fossilise them — demand a correction list after every message.

The clearest win in this course

Language learning is where these tools deliver the most, and it is worth saying plainly after several lessons of caution. A patient conversation partner, available at any hour, who will speak at your level, never gets bored, never judges your accent, and will explain the grammar of the sentence you just got wrong — this did not exist before, at any price, for most learners.

The things that work:

- **Conversation at a controlled level.** *Talk to me in Spanish at roughly A2. Correct me only at the end of each of my messages, in English, in one line.*
- **Comprehensible input on demand.** A short text about a topic you care about, at your level, is the single best thing for acquisition, and it has always been the hardest thing to find. Now you can generate it endlessly.
- **Explaining why**, which textbooks compress. Why the subjunctive here and not there. Why this particle. Why the word order changed.
- **Drilling on the pattern you personally keep getting wrong**, rather than a generic exercise book.
- **Reading support.** Paste a paragraph, get a gloss, then re-read the original.

Where it fails, and the failure is specific

Low-resource languages. Quality tracks the amount of training text. Spanish, French, German, Mandarin, Japanese: strong. Many South Asian, African, indigenous and minority languages: much weaker, sometimes confidently wrong on basic grammar, and prone to producing text that is fluent-sounding and not idiomatic. If you are learning a language with a small web presence, verify with a human speaker or a real textbook. This is one of the sharper inequalities in the technology and it is not improving as fast as the headline capabilities.

Register and dialect. Models default to a standard formal register. They will happily produce textbook Arabic to someone who needs Egyptian, or a formality level that would be strange between friends. Ask explicitly for the variety and register you need, and be aware it may not have it.

Recency in slang and usage. By definition out of date.

Pronunciation. Text feedback on pronunciation is close to useless. Use audio: **Forvo** for individual words by native speakers, and speech tools that give real audio. Many models now support voice conversation, which helps considerably.

Fossilisation: the risk worth naming

In second-language research, **fossilisation** is an error that becomes permanent — you have said it wrong so many times that it now feels correct. It happens to learners who produce a great deal without correction.

An AI conversation partner produces exactly that condition unless you set it up otherwise, because assistants are tuned to be agreeable and will often understand and continue past an error rather than flag it. You will have a pleasant fluent conversation full of mistakes nobody mentioned.

The fix is one clause, used every time:

> *After each of my messages, before replying, list any errors I made and the correct form. Keep the list short. Then continue the conversation naturally.*

And once a week, ask: *what mistakes have I made repeatedly in this conversation?* Those are the ones about to fossilise, and they are the ones worth drilling.

Translation, and what it costs you

Translating your thoughts from your first language into the target one is how you build production ability. Machine translation removes that work entirely, which is why heavy translation use produces learners who can read comfortably and cannot speak.

A defensible arrangement: write it yourself first, then compare with a translation, then note the differences. The comparison is where the learning is. Going straight to the translation is where it is not.

The honest exception: when translation is not the goal — you need to read a paper in another language for your dissertation, or communicate something urgent — use the tool and do not feel bad about it. **DeepL** and Google Translate both have free tiers and are excellent for major language pairs.

Free tools worth the space

- **Anki** for vocabulary, with your own cards, and audio pulled from Forvo.
- **Language Reactor** or similar for subtitles, for input from material you actually want to watch.
- **Tatoeba**, a free database of example sentences with translations, human-produced.
- **Wiktionary**, which is far better than its reputation for etymology and inflection tables.
- A human. Language exchange partners cost nothing and provide the one thing no model does: a person whose good opinion you want, who will be confused when you are unclear, and whose confusion is real.

For students studying in a second language

A large share of students worldwide study in a language that is not their first. Using a model to check your English, understand a lecture, or rephrase a confusing question is entirely legitimate and levels a genuine unfairness.

Two cautions. First, some institutions treat AI-assisted language editing as needing disclosure; check your policy and disclose if in doubt, which costs nothing. Second, detectors flag non-native writing at higher rates than native writing — a well-documented and serious problem covered in Module 7. Keep your drafts and version history. That is the protection, and you should have it whether or not you use any tool.

BEFORE YOU MOVE ON

A learner has a fluent 40-minute conversation in French with an assistant and receives no corrections. Why is this a problem rather than a sign of progress?

A. Because conversation without a native speaker cannot improve production, only comprehension.

B. Because the model was probably operating above the learner's level, so the fluency was the model's rather than theirs.

C. Because uncorrected repeated errors can fossilise, and an agreeable partner will understand and continue past them.

D. Because forty minutes of unstructured conversation is less efficient than the same time spent on spaced vocabulary drilling.

36 · 9 min

Statistics for coursework

THE ONE THING TO KEEP

Give the model the design, the measurement level, the sample sizes and the real question, then run the test in jamovi or R and plot the data before you trust any number, because most analysis errors are visible in the plot.

Choosing the test is not the same as running it

Statistics is where students hand over the most and understand the least, because the software produces a number and the number looks like an answer. A model will suggest a test in one line, and the suggestion will often be reasonable and occasionally wrong in a way that invalidates the whole analysis.

The division of labour that works: use the model to understand *why* a test suits your design, run the test in a deterministic tool, and interpret the output yourself.

The four things that determine the test

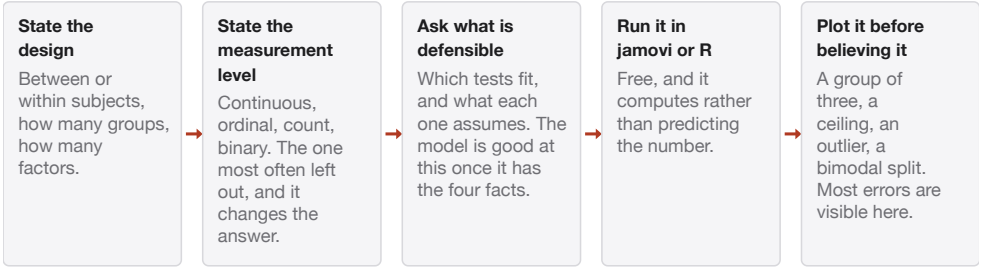
Models give bad statistical advice mainly because students give them incomplete descriptions. Supply all four, every time:

1. **The design.** Between-subjects, within-subjects, paired, repeated measures, how many groups, how many factors.
2. **The measurement level of the outcome.** Continuous, ordinal, count, binary, time-to-event. This is the one most often omitted and it changes the answer completely — a five-point questionnaire item is ordinal, and a t-test on it is a choice that needs defending.
3. **The sample sizes,** including how unequal they are.
4. **The actual question.** Difference, association, prediction, agreement. These need different tests and students frequently describe one while

wanting another.

With those four supplied, the recommendation is usually sound. With any of them missing, the model fills the gap with a default, and defaults are how you end up running a t-test on ordinal data with n of 7.

Choosing the test is not the same as running it



Supply the design, the measurement level, the sample sizes and the actual question, and the recommendation is usually sound. Leave one out and the model fills the gap with a default, which is how a t-test ends up on a five-point questionnaire item with seven people in a group. jamovi and JASP are free and produce the output a marker expects.

Free software that a department will accept

- **jamovi** — free, open source, a proper graphical interface, output that looks like the SPSS output your marker expects. This is the recommendation for most undergraduates.
- **JASP** — free, similar, with unusually good Bayesian options.
- **R** with RStudio — free, what most of the field is moving to, steeper but worth it if you will do more than one project.
- **Python** with pandas, statsmodels and scipy — free, and the right choice if you already program.
- **PSPP** — free, deliberately SPSS-compatible.
- **LibreOffice Calc** or Google Sheets for descriptive statistics and simple tests.

SPSS and Stata are what job ads name, and your university probably has a licence. jamovi and JASP teach the same concepts and produce publishable output, and the transfer takes an afternoon.

The assumptions are the part that gets skipped

Every test has conditions and the marks are usually in checking them.

- **Normality** — of the residuals, not the raw data, which is a distinction models sometimes get wrong. Check with a plot, not only a test; a significance test for normality is oversensitive at large n and underpowered at small n, which is the opposite of useful.
- **Homogeneity of variance** — and know what your test does when it fails, which is usually Welch's correction, often the better default anyway.
- **Independence** — the one that quietly invalidates analyses. Repeated measurements on the same person, students within classes, sites within regions. If observations are clustered, an analysis that ignores it is wrong regardless of what the output says.

Ask the model to *list the assumptions and how to check each one in jamovi*. That is a good, checkable, well-specified request.

Interpreting output without fooling yourself

The standard errors, all still common, all still marked:

- **A p-value is not the probability the hypothesis is true**, and not the probability the result was chance. It is the probability of data at least this extreme if the null were true.
- **Not significant does not mean no effect.** With a small sample it usually means you could not tell.
- **Statistically significant does not mean important.** With n of 5,000 a trivial difference is significant.
- **Always report the effect size and a confidence interval**, not only p. This is now expected in most disciplines and is frequently where the marks are.
- **Multiple comparisons inflate false positives.** Twenty tests at 0.05 give you one significant result by chance on average. Say how many tests you ran.

A model explains all of these well, because they are well represented in the training text. Use it for that.

The move that catches most errors

Before running anything, write one sentence: *I expect X to be larger than Y by roughly Z, because…* Then look at the descriptive statistics and the plot before the test.

Most analysis errors are visible in the plot. A group of 3. A ceiling effect where everyone scored maximum. An outlier at ten times everything else. A bimodal distribution that means your single group is really two. A test run on data you have not plotted is a number produced about a dataset nobody has looked at.

And the honest limitation

Statistics is the area where a confident wrong recommendation is hardest for a beginner to detect, because the output arrives formatted and authoritative either way. If the analysis matters — a dissertation, anything that will be published — take your plan to a human who knows statistics before you collect data, not after. Most universities have a statistics support service, it is free, and it is enormously underused.

BEFORE YOU MOVE ON

A student describes their study as "comparing two groups" and is advised to run an independent t-test. The outcome is a single five-point agreement item and the groups are 9 and 34. What was the root problem?

A. The model gave incorrect statistical advice, which is a known weakness on questions involving assumptions.

B. The description omitted measurement level and sample sizes, so a continuous-data default was applied.

C. A t-test is never valid on questionnaire data, so any recommendation involving one was wrong from the start.

D. The sample sizes were too small for any inferential test at all, so the correct advice would have been to collect considerably more data first.

37 · 8 min

The essay subjects, and what a marker is reading for

THE ONE THING TO KEEP

Generated essays fail identically across subjects — uncontestable thesis, generic evidence, straw objection, reorderable paragraphs — so own the outline, and check every case, statute or named source against a free primary database.

The common structure under law, philosophy, politics, sociology

Essay subjects differ in content and agree almost entirely on what earns marks. A good essay makes a contestable claim, defends it with evidence somebody could check, deals with the strongest objection to it honestly, and does so in a structure where each paragraph moves the argument forward.

Generated essays fail on all four in a recognisable way, and it is worth knowing the failure precisely, because it tells you exactly which parts of your own work you must own.

The claim is not contestable. Generated theses tend towards the balanced and the safe: *the causes were both economic and political*, *the answer depends on context*. True, unarguable, and unmarkable. A marker cannot reward a position nobody would dispute.

The evidence is generic. Named sources without page numbers, examples chosen because they are the famous ones, no engagement with a specific passage or dataset.

The objection is a straw version. *Some might argue X, however this overlooks Y* — where X is stated so weakly that dismissing it costs nothing. Genuine objection handling states the other side at its strongest, which is uncomfortable, which is why it is rewarded.

The structure is smooth and static. Paragraphs of even length, each on a sub-topic, none of which depends on the one before. A real argument has paragraphs that could not be reordered.

Where it helps, precisely

- **Sharpening a thesis.** *Here is my claim. Make it more contestable without making it indefensible. What is the strongest version somebody could disagree with?*
- **Stress-testing.** The sceptic prompt, in a fresh chat. This is the single best use in essay subjects.
- **Mapping the debate** before you read, so you know which positions exist. Then read the actual sources; every name it gives you is unverified.
- **Structural diagnosis of your draft.** *Which paragraph does not advance the argument? Where is my evidence thinnest? Which sentence is doing more work than it can support? Do not rewrite.*
- **Explaining a difficult primary text**, grounded on the passage.

The legal-subject caution

Law students have a specific hazard. Fabricated case citations are the highest-profile AI failure in the whole professional world — lawyers in several countries have been sanctioned for filing briefs with invented cases, complete with plausible names, reporters and paragraph numbers.

And law is jurisdictional. A confident statement of the law is meaningless without knowing whether it is describing your jurisdiction, and models blend them constantly. A rule from one country's statute presented as general principle will be wrong for you in ways that are not visible from the text.

Every case, statute and section gets checked against the primary source. Free sources exist: **BAILII** for UK and Irish material, **CourtListener** and Google Scholar's case law for US, **Indian Kanoon** for India, **AustLII** for Australia, **CanLII** for Canada, and official legislation sites in most countries. There is no excuse for an unverified citation and no defence available if one is found.

Nothing here is legal advice, and the same rule applies to your own life: a model's account of your rights is a starting point for a question to a real adviser, never an answer.

Philosophy, and the risk of plausible argument

Philosophy is unusual: the subject matter is argument itself, so generated text that *resembles* argument is maximally camouflaged. A generated paragraph will move confidently from premise to conclusion with a step that does not follow, and it will read like philosophy because philosophy reads like that.

The defence is the one from the proofs lesson, applied to prose: set out the argument as numbered premises and a conclusion, then ask what licenses each move. Also, philosophers are precise about what a named position actually claims, and models blur positions towards their popular versions. Check the primary text — most canonical philosophy is out of copyright and free at Project Gutenberg or the **Stanford Encyclopedia of Philosophy**, which is free, peer-reviewed and written by specialists, and is a better first stop than any assistant.

The thing that cannot be delegated

Every essay subject rewards the same rare thing: a specific, defensible claim that this particular student arrived at by engaging with particular evidence. There is no way to obtain that from a system that produces the average of what has been written. The average is exactly what a mid-band essay already is.

The practical consequence is that the outline is the essay. If the outline — the claim, the order of moves, the evidence chosen — is yours, everything downstream can be assisted and the work is still yours. If the outline came from a model, no amount of rewriting makes it yours, and it will read like the average, because it is.

BEFORE YOU MOVE ON

Why does a balanced thesis such as "the causes were both economic and political" score badly even when it is accurate?

A. Because it is uncontestable, so there is nothing for the essay to defend and nothing for the marker to reward.

B. Because markers prefer strong claims, and confidence is rewarded over accuracy in essay assessment.

C. Because it is a generated formulation, and markers recognise the pattern from AI-written work.

D. Because it fails to specify which cause was more important, and every essay question implicitly asks for a ranking of factors.

CASE STUDY · MODULE 4

Thirty-four participants and a ceiling

A final-year physiotherapy project in Nagpur, eleven days from submission, with the supervisor on leave and the statistics service booked out.

Meenal Joshi's project asked whether a six-week home exercise programme improved balance in adults over sixty attending an outpatient clinic. She had thirty-four participants, seventeen in each group, and an outcome measured on a fourteen-item balance scale scored from zero to fifty-six.

She described the design to an assistant as comparing two groups before and after. It recommended an independent-samples t-test on the change scores, which is a defensible default given what it was told. She ran it in jamovi, which is free, and got a p-value of 0.31.

Eleven days. Supervisor on leave for another fortnight. The university's statistics support service was showing no appointments until after her deadline.

The first decision was whether that was the end of it. She could report the test she had run, write a discussion around a non-significant result, and submit. Her project is marked on design, execution and interpretation rather than on whether the intervention worked, so a clean negative result written up honestly is a perfectly respectable submission. It would have taken three days and she had eleven.

The alternative was to go back to a question she had skipped, which was whether the analysis matched the data, and that meant plotting it. She had never plotted it. The number had arrived formatted and authoritative and she had treated it as the answer.

She plotted it and found the thing that decided the project. Eleven of the seventeen participants in the intervention group had scored fifty-four, fifty-five or fifty-six at baseline. They were at or within two points of the top of the scale before the programme started. There was no room above them for an improvement to be recorded in. Her instrument could not have detected the effect she was looking for in most of her intervention group, whatever the programme did.

That produced a second and harder decision. She could drop the eleven and analyse the twenty-three who had room to move, which would probably have produced something. It would also have been a decision about which participants to include, made after seeing which ones were inconvenient, which is a different act from the same decision made in advance.

Or she could report the ceiling as the finding: publish the plot, state plainly that the sample was inappropriate for the instrument, give the effect size and its confidence interval, and say that this design could not have answered the question it was set. That meant her project's headline was that she could not tell, and she was convinced it would be marked down for it.

Before deciding she went back to the assistant and asked properly this time: the design, between-subjects with a pre-post measure; the measurement level, an ordinal sum score with a hard upper bound; the sample sizes; the actual question; and the ceiling she had found. The recommendation that came back was a list of options with what each one assumed, which is a good, checkable, well-specified answer to a good, well-specified question. One of its suggestions was to compare her response rate against a published trial's, which she recognised as invalid — two separate studies with different populations and measures are not a comparison — and discarded.

STOP HERE

Before turning the page: what would you have done, and what would it have cost if you were wrong? Write it down. The point of the next section is to compare.

WHAT HAPPENED

She reported the ceiling. The plot went into the results as the first figure, followed by the effect size with its confidence interval, a paragraph on why the original t-test had been the wrong analysis and how she had arrived at it, and a proposal for a more sensitive measure in a sample selected to have room to improve.

The project scored in the top band. The external examiner's written comment quoted the paragraph about the ceiling and described it as the strongest thing in the report. Her own view is that this was partly luck and that a different examiner might have marked the absent result rather than the diagnosis.

Two things she found out too late. Her department's statistics service, which is free, keeps a forty-eight-hour slot for urgent project queries, and nobody in her year knew it existed. And the point at which she needed it was week two, before she collected anything, not week eleven when the data was already fixed. An analysis plan written before collection would have shown the ceiling problem in an afternoon and she could have changed the measure.

She also never got to answer the clinical question she cared about, which was the whole reason she chose the project. She minds about that, and no mark compensates for it.

The first recommendation of a t-test was not stupid. Which of the four descriptors had she left out, and what would supplying it have changed?

She gave the design in outline and left out the measurement level of the outcome, which is the one most often omitted and the one that changes the answer most. A fourteen-item ordinal sum score with a hard upper bound is not the continuous, unbounded quantity a t-test on change scores assumes, and the bound is what produced the ceiling. Supplying it would have prompted a discussion of ordinal outcomes, of what happens near the limits of a scale, and almost certainly of plotting the distribution before choosing anything. The general lesson is that models give bad statistical advice mainly because they are given incomplete descriptions: design, measurement level, sample sizes and the actual question, every time, and the recommendation is usually sound.

Why does 'not significant' not mean 'no effect' here? Say precisely what this design could not have detected.

A non-significant result means the data were compatible with the null, not that the null is true, and with a small sample it usually means you could not tell. Here the reason is sharper than sample size. Eleven of seventeen participants in the intervention group began within two points of the scale's maximum, so even a real and clinically meaningful improvement in their balance had nowhere to be recorded. The instrument was saturated. The study could not have detected an effect in most

of the group it was supposed to detect it in, so the p-value is a statement about a measurement that was never capable of answering the question. That is why the honest report is the plot and the explanation, not the number.

She considered dropping the eleven ceiling participants. Why is that decision different in kind after seeing the data from the same decision made before collection?

Before collection it is an eligibility criterion: a stated rule, applied blind, that anybody could check and that constrains which participants she recruits. After collection it is a choice among analyses, made with knowledge of which choice produces the result she wants, and there is no way for a reader to distinguish it from having tried several and reported the one that worked. The same arithmetic carries a completely different weight depending on when the decision was taken, which is why analysis plans are written in advance. Reporting the ceiling keeps the whole sample, states the limitation, and leaves the reader able to judge. It also happens to be the version she can defend in a viva, where the first question would have been why the eleven are missing.

MODULE 4 · TEST

Check what stuck

6 questions, one right answer each. This one is for you, not for a record: answers and the reason behind each are at the back. If two questions from the same module go wrong, re-read that module before the exam rather than after.

1 Why does a model's arithmetic get worse as the numbers get longer, while its tone does not change?

A. It rounds intermediate results to save tokens in the window
B. Its training data contained mostly two-digit arithmetic examples
C. Longer numbers exceed the context window and get truncated
D. Digits are tokenised without respect to place value

2 What does a generated proof typically look like when it fails?

A. It reaches a conclusion that is obviously false from the start
B. Six correct lines, one unjustified move, the right answer
C. It uses notation from a different branch of mathematics entirely
D. It hedges every step and never commits to a conclusion

3 Why should you write the tests yourself rather than having them generated alongside the code?

A. Tests written by a model take longer to run on large inputs
B. A model cannot express assertions in a testing framework
C. Tests written with the code check it against itself
D. Model-written tests are prohibited under most course policies

4 You carry the units through a generated solution and joules appear where you expected joules per second. What have you learned?

A. The error is above that line
B. The answer needs rounding to fewer significant figures
C. The formula is right and the arithmetic slipped somewhere
D. The units were written in the wrong system and need converting

5 You phrase-search a quotation in Google Books with a date filter and its earliest appearance in print is 1987. What does that establish about an attribution to an 1850s figure?

A. It shows how often historians have repeated the passage since

B. It confirms the speaker did hold the view the line expresses

C. It locates the archive box the original document is filed in

D. A line first printed in 1987 was not said in the 1850s

6 A student says 'I am comparing two groups' and is recommended a t-test. Which omission most often makes that recommendation wrong?

A. The name of the software the department expects to see

B. The measurement level of the outcome

C. Whether the study has already received ethical approval

D. How long the data collection period ran for in total

MODULE 5

Research, sources and evidence

How a citation gets invented, how to find the real paper for nothing, how to read one quickly, and how to say what level of evidence a claim actually rests on.

> BY THE END OF THIS MODULE YOU CAN
>
> Take a claim from a model to a verifiable primary source using free tools, judge what strength of evidence that source carries, and build a bibliography in which you have opened every item

38 · 8 min

Research, And Sources That Exist

> THE ONE THING TO KEEP
>
> Never cite a source you have not opened. Asking the model if it is real does not count.

The failure that keeps happening

In May 2023, a New York lawyer named Steven Schwartz filed a court brief containing six cited cases. Names, docket numbers, judges, quoted passages. All six were invented by ChatGPT.

The part worth memorising is what he did when opposing counsel could not find them. He went back to the chatbot and asked whether the cases were real. It said yes. He asked for confirmation. It produced full text. He filed that too. The judge fined him and his colleague $5,000.

Students do the smaller version of this every term, and it is far easier to catch than AI-written prose, because a marker who wants to check a reference simply checks it. A fabricated bibliography is one of the few genuinely reliable signals a tired academic has.

Why citations in particular

A language model, absent a search tool, is producing statistically plausible text. A citation is an unusually learnable format: author surname, year, a title made of field-appropriate nouns, a journal that publishes in that field, a volume number, a page range, a DOI-shaped string. Every component is drawn from a pattern the model has seen thousands of times.

The result is a reference that is right in every respect except existing.

This is not the model lying. There is no database being consulted and no lookup being fumbled. Unless the tool is actually searching — and you should always know whether yours is — nothing in the process ever checks reality.

The rule, and how to satisfy it

Never cite a source you have not opened. Not summarised, not confirmed. Opened.

How to check, in order, ninety seconds each:

1. **Search the exact title** in Google Scholar, your library catalogue, or your discipline's index — PubMed, JSTOR, SSRN, arXiv. If the exact title returns nothing anywhere, stop. It does not exist.
2. **Resolve the DOI.** Paste it after `https://doi.org/`. A real DOI lands on the paper. An invented one returns an error. This takes five seconds and catches most fabrications.
3. **Check author, year, journal and volume together.** Fabrications frequently attach real, well-known researchers to a paper they never wrote. Finding that Bjork is a real memory researcher tells you nothing about whether this particular Bjork paper exists.

4. **Open it and find your sentence.** This catches the more common failure, which is not invention at all: a real paper cited for a claim it does not make, or makes about a different population, or explicitly argues against.

That last one is worth dwelling on. Genuine papers get miscited more often than fake ones get invented, and miscitation is the error that survives all the way into your finished work.

Search-enabled tools change the risk, not the rule

More tools now retrieve and link — ChatGPT with browsing, Perplexity, Gemini, Claude with search, Copilot. A link you can click is genuinely different from a fabricated reference, and this is real progress.

But the summary can still drift from the source. The tool retrieves a page, compresses it, and the compression may sharpen a hedged finding into a firm one, or lose the sample size, or attribute a result to the wrong study among several on the page. Click through. Read the paragraph. It is quick, and it is the difference between citing and gesturing.

Also notice what the retrieved sources actually are. A confident answer built on three blog posts and a press release is not the same as one built on the paper.

Where AI genuinely helps with research

Use it for the map, not the territory.

```
What are the main competing positions on whether minimum
wage increases reduce employment? Who argues each side,
and what is the strongest objection to each?
```

```
I'm researching informal waste collection in Indian cities.
What search terms would a subject librarian use that I
probably wouldn't think of?
```

```
Here's the methods section of a paper. Explain the
regression discontinuity design in plain terms so I can
follow the results.
```

All three are things it is good at, and none produce a citation. You get orientation, vocabulary, and comprehension — then you go and read.

One more, which will save you an afternoon: after you have your source list, ask it what you appear to have missed, or which viewpoint is absent from your reading. A bibliography that only contains people who agree with you is a common, avoidable mark deduction.

BEFORE YOU MOVE ON

Your draft cites a paper with a plausible title, a real journal, volume and page numbers, and a DOI. What actually establishes that it exists?

A. Asking the model to confirm the reference is genuine and to supply the abstract.

B. The presence of a well-formed DOI, since DOIs are issued by registries and cannot simply be made up.

C. Confirming the listed authors are real researchers who publish in that field.

D. Resolving the DOI or locating the record in a library or index, then opening it and finding the sentence you say it supports.

39 · 8 min

How a citation gets invented

THE ONE THING TO KEEP

Citations are a rigid form that a predictor reproduces easily, so plausible fabrications and real-paper hybrids both occur — resolve the DOI to check existence, and read the page to check support.

Watch it happen

A citation has a rigid form: authors, year, title, journal, volume, issue, pages, DOI. Rigid forms are exactly what a next-token predictor reproduces most easily. Every component is drawn from a strong pattern:

- **Authors** who publish in that area, in the right initials-and-surname format.
- **A year** in the plausible range for that literature.
- **A title** assembled from the vocabulary of the field, often a very good title.
- **A journal** that genuinely publishes that subject.
- **A volume and page range** consistent with that journal's numbering.
- **A DOI** with a real prefix, since prefixes are publisher-specific and heavily represented, and a suffix that is invented.

Every part is individually plausible. The combination has never existed. This is not a bug in the citation feature — there is no citation feature. It is the same process that writes the prose, applied to a highly structured string.

The hybrid, which is worse

Purely invented references are the easy case; a search finds nothing and you delete it. The dangerous ones are hybrids:

- A **real paper with a wrong year**, which breaks your reference list and looks like carelessness.
- **Real authors, real journal, title that does not exist.**
- A **real paper attached to a claim it does not make.** This is the worst, because every element checks out and the citation is still false. Only reading the paper catches it.
- A **real paper whose findings are reversed** — the study found no effect, and the citation is attached to a sentence saying it found one.

The last two are why *the reference exists* is not the same as *the reference supports this*. Two separate checks.

Two separate checks, and only one of them is cheap

Does it exist? Ten seconds.	Does it support the claim? Three minutes.
• Paste the DOI after the doi.org address and press enter	• Get the full text, then find the page your claim rests on
• Search the exact title in quotation marks in Scholar or the library catalogue	• A real paper attached to a claim it does not make passes every other check
• A real DOI lands on the publisher's page; an invented one errors	• So does a real paper whose finding is reversed in your sentence
• Check author, year, journal and volume together, not separately	• Note the page number; that is what makes the citation yours
• Says nothing at all about whether the paper supports your sentence	• If you did not open it, you did not check it

Purely invented references are the easy case; a search finds nothing and you delete the line. The hybrids are the dangerous ones, because every element checks out and the citation is still false. The inference a supervisor draws is not that you used a model. It is that you cited something you never read, which was misconduct long before any of this existed.

Why asking the model does not help

Is this a real paper? is a question about the world, and the model has no channel to the world. It answers from the same process that produced the citation. It will often say yes. Occasionally it will say no about a real paper, which is even more confusing.

This holds when search is switched on, in a modified form. Now it may retrieve something, and the failure becomes a real link attached to a claim the page does not make. Which brings us to the only rule that survives every version of this technology.

The rule

> ***Open it. If you did not open it, you did not check it.***

Open means: reach the actual paper or book, find the page, and read the sentence that supports your claim. Not the abstract if your claim is about a method. Not a search result snippet. The thing itself.

This takes about three minutes per source with the free tools in the next lesson. For a ten-source essay that is half an hour, and it is the half hour that stands between you and the most common academic-misconduct finding in the world right now.

The DOI test, which takes ten seconds

Every real journal article since roughly 2000 has a DOI. Paste it after `https://doi.org/` and press enter. A real DOI resolves to the publisher's page for that article. An invented one gives you a DOI-not-found error.

This single check disposes of most fabricated references instantly. It does not tell you whether the paper supports your claim, only that it exists. Both checks are required and this is the cheap one.

What a supervisor sees

The reason this matters more than it feels like it should: an academic reading your bibliography recognises the literature. They know the journals, often the authors, sometimes the specific paper. A fabricated citation in a field somebody has worked in for twenty years is as visible as a misspelled name in your own language.

And the inference they draw is not *this student used AI*. It is *this student cited something they never read*, which has always been misconduct, long before any of this existed. That is the charge, and it is a fair one.

Using a model for references honestly

There is a legitimate version, and it is a search-strategy conversation rather than a citation request:

> *What are the main research areas relevant to this question, what search terms would find them, and which journals publish this kind of work? Do not give me citations.*

Then search Google Scholar, PubMed or arXiv with those terms yourself. The model is good at knowing the shape of a field and at translating your vague question into the vocabulary a database will match. It is not a bibliography. Used that way, it saves genuine time and generates nothing you then have to un-invent.

BEFORE YOU MOVE ON

A student checks every reference in their draft by resolving each DOI. All resolve to real papers. Why is the bibliography still not verified?

A. Because DOIs can be recycled by publishers, so a resolving DOI may point at a different paper than the one cited.

B. Because resolving a DOI only reaches the publisher's landing page, and a paywall means the full text was never seen.

C. Because the DOI check does not confirm the year, volume and page numbers, which may still be wrong in the reference list.

D. Because existence and support are separate checks, and a real paper can be attached to a claim it never makes.

40 · 9 min

Finding the paper, for nothing

THE ONE THING TO KEEP

Library proxy, Unpaywall, Scholar's all-versions, preprint servers, PMC and emailing the author will reach almost any paper free — and the page number you record while reading is what turns a citation into evidence.

Paywalls are navigable

A great deal of research is behind subscription. Very little of it is genuinely unreachable if you know the routes, and all of the routes below are legal and free.

Your library first. Institutional access covers most of what you need, and the route is usually a proxy login on the library site rather than the publisher's page directly. If you hit a paywall, go through the library catalogue rather than Google, and it often simply opens. Librarians will also request anything the library does not hold, through inter-library loan, usually free for students and usually within days. This service is heavily underused.

Unpaywall. A free browser extension that checks, for any article page, whether a legal open-access copy exists — in a repository, a preprint server, or the author's own site. It finds one surprisingly often. **Open Access Button** does the same job.

Google Scholar. Search the title in quotation marks. Scholar shows links to free versions on the right-hand side, and *All versions* under a result often surfaces a PDF in a university repository.

Preprint servers. arXiv for physics, maths, computer science and increasingly other fields; bioRxiv and medRxiv for life sciences and medicine; SSRN for social science and law; PsyArXiv for psychology. The preprint is frequently the same content as the published paper, minus copy-editing.

PubMed Central, free full text for a large share of biomedical literature, and required for research funded by many public bodies. **Europe PMC** covers similar ground.

Institutional repositories. Most universities require deposit of their staff's papers. Search the author's institution repository directly.

Email the author. This works, it is normal, and researchers are usually pleased. Two sentences: who you are, which paper, that you cannot access it, and would they mind sending a copy. Their email is on the paper.

Books

Internet Archive lends scanned books free, one borrower at a time, and has an enormous collection. **Project Gutenberg** and **Standard Ebooks** for out-of-copyright works. **HathiTrust** for full-text search across scanned books, which lets you find *which page* something is on even when you cannot read the whole book — then use your library for that page. **Google Books** allows preview of many pages and, crucially, full-text search, which is how you verify a quotation.

And the physical library. The book is on a shelf, it is free, and nobody else is looking at it.

Verifying in the right order

A cheap sequence, roughly three minutes per source:

1. Search the exact title in quotation marks. Nothing? Suspect fabrication.
2. Resolve the DOI. Fails? Fabrication or a wrong DOI.
3. Open the abstract. Does it look like the paper your claim needs?
4. Get the full text through one of the routes above.
5. Find the specific sentence, figure or table your claim rests on. Note the page number.

Step five is where the citation becomes yours rather than something you inherited. It is also what makes the writing better, because a page number forces

you to say what the source actually shows rather than what it is generally about.

Search that finds things

Database search is a skill and it is not the same as asking a question. What works:

- **Quotation marks** for exact phrases.
- **Boolean operators** — `AND`, `OR`, `NOT` — supported in most academic databases.
- **Field limits.** `author:`, `title:`, date ranges. Scholar and PubMed both support these.
- **Following citations.** Find one good paper, then look at its reference list for the older work and Scholar's *Cited by* for the newer. This is the most efficient search method there is and it uses somebody else's judgement about relevance.
- **Review articles first.** A recent review gives you the map of a field and a curated reference list. Search for `[topic] systematic review` or `[topic] narrative review`.

A model is genuinely useful for step zero here: turning your question into the vocabulary a database uses. Fields have terms of art, and searching for the plain-English version returns nothing. Ask what the technical terms are, then search for those.

Managing what you find

Zotero is free, open source, and better than the paid alternatives. Browser plugin saves a reference and the PDF in one click, generates bibliographies in any style, and syncs. Install it in your first week and never assemble a reference list by hand again.

One caution that comes up constantly: Zotero's automatic metadata is usually right and occasionally wrong, especially for chapters and older material. Check what it captured. A reference list where every entry is subtly malformed is a different kind of avoidable mark loss.

BEFORE YOU MOVE ON

Why is following a good paper's reference list and its "cited by" list often more efficient than searching a database directly?

A. Because citation chains avoid the paywall problem, since cited papers are more likely to be open access.
B. Because databases index only a fraction of published work, whereas reference lists capture everything in a field.
C. Because reference lists are curated by the paper's authors, so you inherit an expert's judgement about what is relevant.
D. Because it removes the need to know a field's technical vocabulary, which is the main obstacle to effective database search.

41 · 8 min

Reading a paper without reading all of it

THE ONE THING TO KEEP

Read in the order abstract, figures, aim, discussion, limitations — and check any claim that matters against the results section rather than the abstract, because abstracts systematically overstate.

Nobody reads papers front to back

Researchers do not start at the first word and finish at the last. They triage. For an undergraduate reading twelve papers for an essay, the difference between triaging and not is the difference between finishing and not.

The order that works:

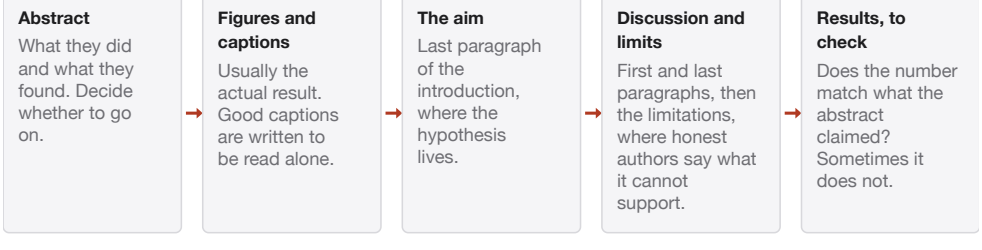
1. **Title and abstract.** What did they do, what did they find. Decide whether to continue. Most papers stop here, and that is a successful outcome, not a failure.

2. **Figures and tables, with their captions.** This is the surprising one. The figures usually contain the actual result, and captions in a good paper are written to be readable alone. You can often learn what a paper found in ninety seconds this way.
3. **The last paragraph of the introduction.** Where the specific aim and hypothesis live.
4. **The first and last paragraphs of the discussion.** What they think it means, and what they concede.
5. **The limitations section.** Read this properly. It is where the honest authors tell you what the paper cannot support, and it is exactly what you need in order to cite it accurately.
6. **Methods**, only if the result matters to you. Then read it slowly, because this is where a study is strong or weak.
7. **Results text**, to check the numbers match what the abstract claimed. They sometimes do not.

Where the model helps and where it must not

The safe pattern: read the paper yourself in that order, and use the model on the parts you cannot follow, grounded on the pasted text.

The order researchers actually read in



Methods come last, and only if the result matters to you. The step almost nobody does is the final one: abstract conclusions have a documented tendency to be stronger than the results section supports, so a claim carrying weight in your essay gets checked against the results, which takes a minute. Stopping after the abstract is a successful outcome, not a failure.

> *Here is the methods section. Explain what a mixed-effects model is doing here and why they would have chosen it over a repeated-measures ANOVA. Use only what is in the text; say what is missing rather than filling it in.*

That is a good use. It is a comprehension aid on material in front of you.

The unsafe pattern is asking for a summary and citing the summary. Two things go wrong. Summaries drop conditions and hedges — the *in older adults only*, the *when baseline was controlled* — and those qualifications are frequently the entire finding. And a summary you did not check gives you no page number and no sentence, so you cannot cite honestly.

The questions to ask of any study

Whatever the field:

- **How many participants or observations, and how were they selected?** A study of 24 undergraduates supports different claims from one of 24,000 adults.
- **Compared with what?** No control group means no causal claim. A comparison with nothing is not a comparison.
- **Was anything randomised?** This is the difference between *associated with* and *caused*.
- **What exactly was measured?** *Wellbeing* measured how? Self-report on one item is not the same as a validated scale.
- **How many analyses were run?** Twenty tests produce a significant result by chance.
- **Who funded it, and who benefits from the finding?** Not disqualifying; relevant.
- **Is this the first report or a replication?** First reports are much more often wrong than the confidence of their abstracts suggests.

A model can be asked to apply this list to a paper you have pasted, and that is a good use — it is applying a checklist to supplied text rather than recalling anything. Verify the answers against the paper.

Abstract inflation

Worth knowing as a specific hazard: abstracts systematically overstate. They are written to be read and cited, they omit the qualifications, and there is a

documented tendency for abstract conclusions to be stronger than the results section supports. In several fields this has been measured, and the mismatch is not rare.

So if a claim matters, check it against the results, not the abstract. This is the single highest-yield habit in reading research, and it takes one minute.

Note-taking that survives

For each paper, four lines in your own words:

> ***Claim.*** *What they say they showed.*

> ***Evidence.*** *Design, n, key number, with the page or figure.*

> ***Limitation.*** *What it cannot support.*

> ***Use.*** *Which part of my argument this is for.*

Written in your own words, because a copied summary is a copying-into-the-essay accident waiting to happen and because paraphrasing is where you find out whether you understood. Zotero holds these notes attached to the reference.

Thirty papers read this way is a working knowledge of a literature. Thirty AI summaries is a folder.

BEFORE YOU MOVE ON

Why does the reading order put figures and captions second, before the introduction and methods?

A. Because figures are the part least likely to be affected by the authors' interpretation of their own data.
B. Because methods sections are too technical to be useful until you already know what the result was, so they must come after everything else.
C. Because the introduction is written last by most authors and therefore reflects the finished argument rather than the study.
D. Because captions are usually written to stand alone, so the actual result is available in about ninety seconds.

42 · 8 min

What counts as evidence

THE ONE THING TO KEEP

Match the claim to the design — causal language requires randomisation, and comparing response rates across two separate trials is invalid however natural it reads.

Not all sources are the same weight

Students lose marks by citing a blog post as though it were a trial, and by citing a trial as though it settled a question. Knowing what a source can support is a skill markers reward heavily, and it takes about one lesson to acquire the outline of.

Roughly ascending, for a claim about whether something works:

- **Anecdote.** One person's report. Establishes that something happened once, and nothing else.

- **Case report or case series.** Documented individual cases. Valuable for the rare and the new; cannot establish frequency or cause.
- **Cross-sectional survey.** A snapshot. Shows association. Cannot establish direction — did the exercise cause the mood, or the mood cause the exercise.
- **Cohort study.** Followed over time. Establishes temporal order, which is a real gain. Still vulnerable to confounding: the people who chose the exposure differ from those who did not, in ways you may not have measured.
- **Randomised controlled trial.** Allocation by chance, so on average the groups differ only in the intervention. This is what licenses causal language. Its weaknesses are its own: often small, often short, often on unrepresentative participants.
- **Systematic review and meta-analysis.** All the eligible studies located by a stated method and combined. The strongest form, and only as good as the studies in it. A meta-analysis of twelve small biased trials is a precise estimate of a biased effect.

The hierarchy is a guide, not a law. A well-run cohort study of 100,000 people beats a badly-run trial of 30. What matters is being able to say why.

What a source can carry, strongest at the top

Systematic review and meta-analysis	All eligible studies, found by a stated method. Only as good as the studies in it.
Randomised controlled trial	Allocation by chance, so the groups differ on average only in the intervention. This licenses causal language.
Cohort study	Followed over time, so temporal order is established. Still open to confounding you never measured.
Cross-sectional survey	A snapshot. Shows association and cannot tell you which way it runs.
Case report or case series	Valuable for the rare and the new. Cannot establish frequency or cause.
Anecdote	Establishes that something happened once, and nothing else.

A guide rather than a law: a well-run cohort of a hundred thousand beats a badly-run trial of thirty, and the mark is for being able to say why. Match the claim to the design. Causes needs randomisation. Is common needs a representative sample. Works better than needs a direct comparison, not two separate studies read side by side, however natural that reads.

Peer review: what it does and does not do

Peer review means two or three people in the field read the manuscript and recommended publication, usually unpaid, usually in a few hours of work. It filters out a lot of obvious rubbish. It does not verify the data, does not usually recompute the analysis, and does not guarantee replication. Plenty of peer-reviewed findings have failed to replicate — the replication crisis in psychology and several other fields is exactly this.

So *peer-reviewed* is a floor, not a certificate.

Preprints have not been reviewed. They are legitimate to cite in many fields, especially fast-moving ones, and you should say so: *(preprint, not peer reviewed)*. This is normal practice, not an admission.

Predatory journals publish anything for a fee and exist to look like real journals. Check whether the journal is in the **Directory of Open Access Journals**, whether the editorial board is real, and whether anyone in your field has heard of it. A model will happily cite a predatory journal without comment.

The specific claims that need specific evidence

Match the claim to the design:

- *X causes Y* needs randomisation, or an extremely well-argued natural experiment.
- *X is associated with Y* needs an observational study, and the sentence must not drift into causal language in the next paragraph, which is where most essays lose the mark.
- *X is common* needs a representative sample, which is a much harder thing to obtain than people assume.
- *X works better than Y* needs a direct comparison, not two separate studies of each.

That last one is worth dwelling on. Comparing across studies — this drug had 60 per cent response in one trial, that one 45 per cent in another — is invalid,

because the populations, measures and conditions differed. A model will do it happily if asked, and it reads perfectly.

Where the model helps

- **Explaining a design** and why it supports what it supports.
- **Naming what would be needed** to establish a claim you want to make. Excellent for planning research.
- **Spotting causal language over correlational evidence** in your own draft: *find every sentence where I claim causation, and tell me what design each cited source used*. A genuinely good self-check.

The honest state of things

Some of what you will cite is contested, and saying so is a strength rather than a hedge. Fields disagree; results reverse; large literatures rest on findings that later failed to replicate. Writing *the evidence here is mixed, with X finding one thing in a large cohort and Y finding the opposite in a smaller trial* is more accurate and better rewarded than picking one and asserting it.

A model will usually give you the consensus position smoothly and flatten the disagreement, because smooth consensus is what most text about a topic looks like. Ask specifically: *what is contested about this, and who disagrees?* Then go and read both sides.

BEFORE YOU MOVE ON

An essay cites two separate trials, one reporting 60 per cent response to drug A and another 45 per cent to drug B, and concludes A is more effective. What is wrong?

A. Neither figure means anything without confidence intervals, so no comparison of any kind can be made.
B. The comparison is across studies with different populations, measures and conditions, so the percentages are not commensurable.
C. The trials were not randomised against each other, so this is observational evidence rather than causal evidence.
D. Response rate is the wrong outcome for a comparison of effectiveness, which requires a head-to-head measure of clinical benefit.

43 · 8 min

When it searches the web

THE ONE THING TO KEEP

Retrieval replaces invented sources with real sources attached to claims they do not make, so the check is unchanged: open the link, find the sentence, and follow reporting back to the original.

A different failure, not no failure

Most assistants now retrieve web pages and cite them. This is a genuine improvement: the model is working from text it just fetched rather than compressed memory, so pure invention drops sharply.

What replaces it is subtler and easier to trust by mistake.

The link is real, the claim is not in it. The most common failure. A citation attached to a sentence the page does not support, sometimes because the

page is about the topic generally and the specific claim was generated around it. Independent evaluations of search-grounded assistants keep finding this in a substantial minority of citations.

The source is weak. Retrieval optimises for what ranks, not for what is reliable. A content-farm article, an SEO page, a five-year-old forum post and a Cochrane review are all just retrieved text.

The summary distorts. Even from a good page, conditions and hedges drop out. *May reduce risk in some populations* becomes *reduces risk*.

Only part was read. Retrieval usually fetches a portion. The model can answer confidently about a document it saw a fragment of.

Circularity. An increasing share of the web is itself model-generated. A claim can be confirmed by three sources that all originate from the same unverified generated text. Multiple agreeing pages are weaker evidence than they used to be.

The check, which has not changed

Open the link. Find the sentence. That is the whole method, and it is why this course keeps repeating it: every generation of this technology changes what fails and leaves the check identical.

When you open it, ask three things:

1. **Does this page say what the answer said?**
2. **Is this page a source or a repeater?** A news article about a study is not the study. Follow it back.
3. **Would I have cited this if I had found it myself?** A page you would not have chosen does not become citable because a machine surfaced it.

Deep research modes

Several tools now offer a longer, multi-step research mode that runs many searches and returns a report with dozens of citations. These are genuinely more thorough, and they concentrate the risk rather than removing it: a twen-

ty-page report with sixty citations invites acceptance precisely because checking it looks expensive.

Use them as a **map**, which is what they are good for: what the areas are, which names recur, which terms to search. Then do your own reading of the sources that matter. Never submit a report of this kind, or its reference list, as your own work — and note that a bibliography assembled by a tool you did not verify is a bibliography of sources you have not read, which is the misconduct finding from two lessons ago.

Wikipedia, used correctly

Wikipedia is an excellent starting point and not a citable source, and both halves are true for the same reason: it summarises, and it tells you what it summarised from.

The correct use is to **read the article and then use the reference list at the bottom.** Those are the real sources; go to them, read them, cite them. The article's talk page is also unusually informative — it is where disputes about the content are argued out, so you can see immediately whether a claim is contested.

And for offline work, **Kiwix** packages Wikipedia for local reading, which is free and useful with intermittent data.

The one place it is unambiguously better

Retrieval is a real improvement for anything time-bound: this year's entry requirements, the current version of a library, a deadline, an exam board specification, who currently holds an office. A model without search has no route to any of these and will answer from whenever its training stopped, often without saying so. If your question has a date in it, or implicitly has one, switch search on and open the page.

Building the habit

A short discipline for any answer with links:

- Open every link you intend to rely on. Not the ones you are ignoring.
- Search within the page for the specific term from the claim.
- If the page is reporting somebody else's work, follow it back to the original.
- If you cannot find the claim in the page, the claim is unsupported. Delete it or find a real source. Do not keep it because it sounded right.

That last instruction is the one people break. A claim that survived because it was plausible, attached to a source that does not contain it, is the single most common way a careful student ends up in an integrity meeting.

BEFORE YOU MOVE ON

Why does the growth of AI-generated web content weaken the practice of confirming a claim across several independent-looking pages?

A. Because several pages may all derive from the same unverified generated text, so agreement is not independence.

B. Because generated pages are more likely to contain factual errors than human-written ones, raising the error rate of any sample.

C. Because generated pages rank lower in search, so retrieval increasingly returns only a narrow slice of the web.

D. Because generated content lacks citations of its own, so the trail back to a primary source is broken at the first step.

44 · 8 min

Notes that survive the term

THE ONE THING TO KEEP

Compression is the work, so an AI summary is a note nobody encoded — write in your own words, keep a confusion section, and use the model to interrogate your notes rather than to produce them.

The test of a note

A note is good if, three months later, it reminds you of something you understood. That is the only criterion, and it rules out most of what students produce: transcriptions, highlighted PDFs, and AI summaries.

The failure of the first two is old. A transcript is a recording of a lecture happening, not of you understanding it. Highlighting marks text as important and produces recognition, which is why it consistently rates among the least effective study methods when tested.

The failure of the third is new and worth naming precisely. **An AI summary is a note nobody encoded.** Compression is the work. A summary you did not write is a summary someone else's process performed, and what you retain from reading it is roughly what you retain from reading anything — a feeling of familiarity, and no structure of your own to hang the material on.

Notes in your own words, always

The rule underneath everything here: **write it as you would say it to a friend.** If a sentence in your notes could have been copied from the source, it probably was, and it will teach you nothing when you return to it.

This also solves a practical problem. Notes containing copied sentences are how accidental plagiarism happens: three weeks later you cannot tell which

sentences were yours. Paraphrasing at note-taking time removes the risk entirely, and it makes the note better. Two problems, one habit.

If you must keep an exact quotation, mark it unmistakably — quotation marks, the page number, and something visual you cannot miss.

A structure that works

For a lecture or a reading, four sections:

- **Question.** What is this trying to answer? One line, written before or immediately after.
- **Core.** The three to five things that would let you reconstruct the rest. Not everything.
- **Confusion.** What you did not follow. This is the most valuable section and the one everyone omits. It becomes next session's opening.
- **Connection.** What this links to elsewhere in the course, or contradicts.

Then, within a day, close the notes and write the summary from memory in three sentences. That is retrieval, it takes two minutes, and it converts a passive record into a study session.

Where the model belongs in note-taking

Not as the writer. As the interrogator.

- **After you write.** Paste your notes: *what have I misunderstood, what is missing, and what would somebody who knew this subject add?* Diagnosis on your work.
- **Turning notes into questions.** *Turn these notes into fifteen questions, a third of which require applying the idea to a new case.* Then answer them from memory.
- **Filling a named gap.** You know what you missed; ask about that specifically, with your notes pasted for context.
- **Cards.** Generating Anki cards from your notes, checked against the source before they enter the deck.

The boundary is straightforward: the model works on notes you wrote. It does not write them.

Recorded lectures and transcripts

Automatic transcription is genuinely useful for accessibility, for second-language students, and for going back to the one thing you missed. **Whisper** runs free and locally, and most platforms now transcribe automatically.

The trap is that a transcript feels like notes and is not. It is the lecture, in text, unprocessed. Use it to check a specific point, or as a source to make notes *from*. Never as the notes themselves — a folder of transcripts at revision time is a folder of lectures you now have to attend again.

Tools, all free

Obsidian — free for personal use, plain markdown files you own, works offline, links between notes. **Joplin** — open source and free, similar, syncs. Plain markdown files in a folder — free, portable, will still open in twenty years. **Anki** for the retrieval layer.

The tool matters much less than people spend time believing. Elaborate note systems become a productivity hobby that substitutes for studying. Plain files, four sections, own words, reviewed by retrieval.

What to keep

At the end of a module you should have three things: your notes, your confusion list with resolutions, and your unaided-check log from Module 3. Those three, together, are a genuine account of what you learned and how. They are also the process record that Module 7 is about — and unlike anything you could assemble after the fact, they exist already, because you made them while working.

BEFORE YOU MOVE ON

Why does paraphrasing at note-taking time solve two problems at once?

A. It shortens the notes and makes them faster to revise from later.

B. It creates a record in your own voice, which serves as evidence of authorship if your work is questioned.

C. It performs the encoding work and removes the risk of copied sentences reappearing in your essay.

D. It forces you to identify the source's argument structure, which is what makes a summary useful three months later and what protects against misattribution.

45 · 7 min

Building a bibliography you can defend

THE ONE THING TO KEEP

Every line of a reference list asserts that the work exists, that you read it, and that it supports the claim — so a model can be cited as a method you used, never as evidence for a fact.

The list is a set of claims

A reference list is not decoration. Every entry asserts: this work exists, I consulted it, and it supports what I attached it to. Three claims per line, and you are answerable for all of them.

That framing settles most questions about how to do this. A reference you did not open makes a false claim, whether it came from a model, a friend's essay, or a paper's own reference list copied without checking.

Zotero, in ten minutes

Zotero is free and open source, and it does the job better than EndNote, which universities pay for. Install the desktop app and the browser connector.

Then the workflow is: read a paper, click the connector, and the reference plus the PDF is saved. Write in Word, LibreOffice or Google Docs with the Zotero plugin, insert citations as you write, and generate the bibliography in whatever style your department demands. Changing style is one dropdown, which matters more than it sounds when a department switches from Harvard to APA between years.

Two cautions. Zotero's automatic metadata is usually right and sometimes wrong — check chapters, older books and anything from an unusual site. And keep your notes in Zotero attached to the item, so the four-line note from the previous lesson lives with the reference rather than in a separate file you will lose.

Free alternatives: **JabRef** if you use LaTeX, **Mendeley** free tier, or a plain BibTeX file maintained by hand, which is entirely respectable.

Why models get citation formatting wrong

Styles are rule systems with many exceptions, and a model reproduces the general pattern while missing the specific case: a chapter in an edited volume, a translated work with two dates, a government report, a dataset, a preprint, a court judgment. The output looks right and is subtly wrong in the way that costs marks in a coursework rubric with a presentation criterion.

Zotero applies the actual style specification, which is maintained by people who care about the edge cases. Use the tool.

One legitimate use of a model here: *explain why this reference is formatted this way in APA* when you cannot work out what your source type is. Explanation, not generation.

Citing an AI tool

If you used a model in a way your institution asks you to declare, cite it in whatever form they specify. Most style guides now have a form for it, and most treat it as a personal communication or software rather than a source, because it is not retrievable — nobody else can go and read what it told you.

That unretrievability is exactly why a model can never be a source for a factual claim. A citation exists so a reader can check. A conversation nobody else can open is not checkable, so it supports nothing.

What you cite, therefore, is the *use* — as a disclosure of method — and never as evidence for a fact. The fact needs a real source underneath it.

Before you submit

A fifteen-minute check that has saved a great many people:

1. Every in-text citation appears in the reference list, and vice versa. Zotero does this; check anyway.
2. Every DOI resolves.
3. For every reference: can you say, out loud, what it argues and which page you used?
4. Any reference where the answer to 3 is no comes out. Not *gets checked later*. Comes out.
5. Every direct quotation matches the source word for word, including punctuation.
6. Nothing is cited from a summary you read rather than the source itself.

If that check removes half your bibliography, that is a finding about the essay, and it is much better to discover it now.

The reference list as a reading plan

One reframing makes all of this less painful. Build the bibliography as you read, from the first day, rather than assembling it the night before. Every paper you read gets saved and annotated at the moment you read it, with the

four-line note and the page numbers. The reference list then writes itself, and it contains exactly the works you actually used.

Students who assemble a bibliography at the end are doing something different and much harder: reconstructing, under time pressure, which source said what. That reconstruction is where fabricated and misattributed citations enter an otherwise honest essay.

The standard, stated plainly

Cite what you read. Read what you cite. Say where in it.

That standard predates every tool discussed in this course by several centuries, and none of them changed it. What changed is only how easy it has become to produce a list that looks like it meets the standard without meeting it.

BEFORE YOU MOVE ON

Why can a conversation with a model never serve as the source for a factual claim in an essay?

A. Because institutions prohibit citing AI tools, so any such reference breaches the assessment rules.

B. Because models are frequently wrong, so anything they assert requires corroboration before use.

C. Because the conversation is not a published work and citation styles have no format that can properly represent an unpublished exchange.

D. Because nobody else can retrieve it, and a citation exists so a reader can check the claim.

46 · 8 min

Misinformation, and what you pass on

THE ONE THING TO KEEP

Verify provenance rather than appearance — reverse image search, the primary record, the date, and lateral reading about who runs the page — because generated media can no longer be judged by looking at it.

The same skill, outside the essay

Everything in this module is source criticism, and it does not stop at the bibliography. The most consequential use of it is what you forward to a family group, repeat in an argument, or believe about a subject you will vote on.

AI has changed this in three specific ways, and it is worth being precise rather than alarmed.

Volume. Producing plausible text and images is now nearly free, so there is more of everything, including more well-written wrong things. The ratio has shifted, not the nature of the problem.

Quality of forgery. Synthetic images, audio and video are good enough that visual inspection is no longer a reliable test. The old advice — look for six fingers, look for the artefacts — is already out of date and will get more so.

Personalisation. Generated content can be produced for a specific audience at scale, so the material you encounter is more likely to be shaped to what you already believe.

What still works

Provenance, not appearance. Since you cannot verify a video by looking at it, verify where it came from.

- **Who published this first?** Reverse image search — Google Images, TinEye, Yandex — will often find an image's earlier appearances, which frequently reveals it is from a different event or year.
- **Does any organisation with something to lose report it?** Not agreement across many sources, which is now cheap to manufacture, but reporting by an organisation that would suffer for being wrong.
- **Is there a primary record?** For a claim about a law, a court case, a statistic or a scientific finding, the primary record is usually online and free. Statistical agencies, official gazettes, court databases, the paper itself.
- **What is the date?** Old material recirculated as new is far more common than fabrication, and much easier to check.

Lateral reading is the technique that research on this consistently finds effective: instead of scrutinising the page in front of you, open new tabs and find out who runs it. Professional fact-checkers do this immediately; students left to themselves usually stay on the page and evaluate its design, which is exactly what a designed page is for.

Detection tools do not settle it

There are tools claiming to detect AI-generated images, audio and text. Their accuracy varies enormously, they degrade as generators improve, and they produce both false positives and false negatives at rates that make a single verdict unusable as evidence. Treat any such output as one weak signal.

Provenance standards like **C2PA**, which attach signed metadata about how a file was made, are more promising because they carry a verifiable chain rather than a guess. Adoption is partial and metadata can be stripped, so absence of provenance proves nothing yet.

Your own exposure

The uncomfortable part. You are more likely to pass on something false when it confirms what you already think, when it makes you angry, and when it is about a group you do not belong to. That is not a character flaw; it is how everybody works, and knowing it is the only available defence.

So the practical rule: **the stronger your reaction, the more checking it needs.** Content that makes you want to share it immediately is content built to do that, whether by a person, an algorithm or a generator.

A worked check, in ninety seconds

A video circulates showing a politician saying something inflammatory. Before forming a view:

1. Reverse image search a frame. If the footage is old or from elsewhere, this usually surfaces it immediately.
2. Search the claimed quotation in quotation marks with a date filter. Real remarks by public figures are reported by many outlets within hours.
3. Look for the full clip. Edited-down quotations that reverse a meaning are far more common than synthetic video, and much cheaper to make.
4. Check who first posted it and when.

Notice that only step one is about the file, and none of it requires judging whether the video looks real.

Why this belongs in a study course

Because it is the same skill, and because your subject is not sealed off from it. A biology student will meet health misinformation. An economics student will meet manipulated charts. A law student will meet confident wrong claims about rights, in enormous volume. Being the person in your family or your group who can find the primary source is a small, real, useful thing, and it comes free with learning to write an essay properly.

And the traffic runs both ways. If you would not put an unverified claim in an essay because a marker would catch it, notice that the standard should not depend on whether anybody is marking.

BEFORE YOU MOVE ON

Why is agreement across many web sources a weaker check than it used to be?

A. Because search engines now rank generated content higher, so agreeing pages are more likely to be surfaced together.

B. Because generating many plausible pages is nearly free, so agreement no longer implies independent sources.

C. Because most people now read summaries rather than original pages, so agreement is measured on derived text.

D. Because fact-checking organisations cannot keep pace with the volume, so incorrect claims stay uncorrected across more sites for longer.

CASE STUDY · MODULE 5

Five authorities nobody had opened

A three-person moot court team from a law college in Kochi, thirty-six hours before a national competition memorial deadline.

The memorial cited fourteen authorities. Nine of them Aravind Nair and Shruti Pillai had found themselves, in the library and on Indian Kanoon, over five weeks of evenings. The other five had come from Fahad, their researcher, who had run a deep research mode on an assistant and got back a nineteen-page brief with more than forty citations in about eleven minutes.

The brief was good, and it is important to the story that it was good. It was better organised than anything the three of them had produced. It named two lines of argument they had not thought to look for, and it used the technical vocabulary of the area correctly, which had saved Fahad a week of learning what to search for. Two of its authorities were now load-bearing for the argument the team was proudest of.

Thirty-six hours before submission, Aravind was drafting the oral outline and hit something he could not talk himself out of. For five of the fourteen he could not say which paragraph he was relying on, because he had never opened the judgment. Neither had Shruti. Fahad had read the report.

The rule that applied had nothing to do with AI and predated it by a long way. Every line of a citation asserts three things: that the work exists, that you consulted it, and that it supports what you attached it to. They were making three claims about five authorities and could stand behind one of them.

Every option cost something real, which is why it took them two hours to decide.

Checking all fourteen properly on Indian Kanoon and through the library's subscription database would take the evening. That evening was already allocated to oral rehearsal, and orals carry more of the competition score than the written memorial does. Rehearsal is also the part that cannot be done at two in the morning, because it requires somebody to sit opposite you and interrupt.

There was the cost to the team. Raising it meant telling Fahad that five weeks of his work needed auditing, and Fahad was the person who would have to do most of the auditing. He had not been dishonest and nobody thought he had. He had used a tool that produces a report, and a report with forty citations in it looks exactly like a bibliography.

There was the cost to the argument. If either of the two load-bearing authorities failed, a section written over three weeks came out of the memorial at thirty-six hours, and the argument it supported was the one they had built the oral round around.

And there was the option of submitting as it stood. Realistically, nothing might have happened. Written memorials are marked under time pressure by people with sixty of them, and nobody checks fourteen citations. Against that: the bench in the oral rounds included two practising advocates, a single question about a paragraph number would have been enough, and the competition rules that year made misrepresentation of authority a ground for disqualification of the whole team.

Aravind put it to the other two at seven in the evening. They checked until two in the morning, three people, two laptops and the library's remote login, taking one authority each and reading the judgment rather than the headnote.

STOP HERE

Before turning the page: what would you have done, and what would it have cost if you were wrong? Write it down. The point of the next section is to compare.

WHAT HAPPENED

Of the five, two existed, said what the brief said they said, and went into the memorial with paragraph numbers beside them.

The third existed and was cited for paragraph sixty-three. The judgment runs to forty-one paragraphs.

The fourth existed, the quoted passage was real, and it came from the dissenting opinion. The brief had presented it as the holding. Every element of that citation checks out — real case, real reporter, real volume, real words on a real page — and the proposition it was offered for is the opposite of what the court decided.

The fifth did not exist. The party names were a plausible pairing of two real litigants from unrelated matters, the reporter and the year were real, and there is no such case.

They dropped the argument that rested on the dissent and on the case that did not exist, and rewrote the section around the authorities that survived. The written memorial scored sixty-two against a round average of sixty-eight, and they believe the missing section is why. In the second oral round a judge asked Shruti for the paragraph in one of the authorities they had kept, and she had it, and the exchange lasted forty seconds instead of ending the round.

They will never know what would have happened had they submitted all five unchecked. Probably nothing. That is not the same as it having been a reasonable risk.

The rule the team adopted afterwards is one line in the shared document: whoever cites it opens it, and writes the paragraph number beside it, and no citation enters the draft without one. Fahad wrote it.

Two of the four failures would have passed an existence check. Identify them, and state the second check and why it is the expensive one.

The dissent presented as the holding, and the real judgment cited for a paragraph that does not exist. Both have a real case, a real reporter and a real citation, so any check that asks does this exist returns yes. The second check asks whether the source supports the claim, and it can only be answered by getting the full text, finding the passage, and reading it — which is three minutes a source rather than ten seconds, and which is why it gets skipped. It is also the check that matters more, because genuine works get miscited far more often than fake ones get invented, and a miscitation survives every cheap test all the way into the finished document. The two checks are separate and both are required.

The deep research report was more thorough than Fahad could have been in eleven minutes. Where exactly did using it go wrong?

Not in running it, and not in reading it. These modes are genuinely better at the thing they are good at, which is producing a map: what the areas are, which names recur, which terms to search, which lines of argument exist. The error was treating the map as the territory, by letting the report's reference list become the team's reference list. A long report with forty citations invites acceptance precisely because checking it looks expensive, so the risk is concentrated rather than removed. The correct use was to take the report's terms and case names, search them on Indian Kanoon, read the judgments, and build the authorities from what he had read — which would have taken days rather than eleven minutes, and would have produced nine good authorities instead of five unopened ones.

Fahad had not lied about anything. What charge would still have been available if the memorial had gone in unchecked, and why does it predate all of this?

Citing sources he had not read. That is what a reference list asserts and what he could not have stood behind, and it has been misconduct for as long as there have been footnotes — it is the same offence as copying a citation out of somebody else's bibliography without opening the work, which students did long before any of this existed. The inference an examiner draws on finding a fabricated or misapplied authority is not that a model was used. It is that the person cited something they never opened, which is a fair charge and one that intention does not soften. That is why the rule the team wrote afterwards says nothing about AI at all.

MODULE 5 · TEST

Check what stuck

6 questions, one right answer each. This one is for you, not for a record: answers and the reason behind each are at the back. If two questions from the same module go wrong, re-read that module before the exam rather than after.

1 You paste a DOI after doi.org and it resolves to the publisher's page. What has that settled?

A. That the paper supports the sentence you attached it to
B. That the authors listed actually wrote the paper together
C. That it exists, and nothing about support
D. That the journal is indexed and is not a predatory title

2 An assistant with web search attached returns answers with clickable links. What has changed about the failure?

A. Real links appear, attached to claims the pages do not make
B. The model stops summarising and quotes pages word for word
C. Citations become reliable, so opening the link is unnecessary
D. Every retrieved page has been reviewed before it is cited

3 Why check a claim against a paper's results section rather than its abstract?

A. Abstracts are written last and often omit the method entirely
B. Abstracts overstate, so check the claim against the results
C. Abstracts are written by editors rather than by the authors
D. Abstracts are the only part of a paper that is peer reviewed

4 Which feature of a study is what licenses a sentence to say that X caused Y?

A. A large cross-sectional survey with a representative sample
B. A cohort study that follows the same people over several years
C. Two separate trials whose response rates can be compared
D. Allocation by chance between the groups

5 Why is an AI summary of a chapter a poor substitute for notes you wrote?

A. Compression is the work, and somebody else performed it

B. Summaries are usually too short to cover a whole chapter

C. Copying a summary into your notes is a copyright breach

D. The model cannot read a lecture slide accurately enough

6 In what capacity can a language model legitimately appear in your work?

A. As a secondary source, cited the way a textbook would be

B. As a primary source when it is the only place a claim appears

C. As a method you used, never as evidence for a fact

D. As a personal communication, which makes the fact checkable

MODULE 6

Writing, and keeping your voice

Where in the writing process the thinking actually happens, how to get help at the points that do not displace it, and what makes prose recognisably yours.

> BY THE END OF THIS MODULE YOU CAN
>
> Produce written work whose argument, structure and sentences are yours, use assistance at the specific points that do not displace your thinking, and show a process record demonstrating how the piece was made

47 · 8 min

Writing With It Without Losing Your Voice

> THE ONE THING TO KEEP
>
> If the first draft is not yours, everything after is editing someone else's argument.

The order of operations decides everything

There is one structural decision in AI-assisted writing, and it happens before the first sentence.

If the model drafts and you edit, the argument is already made. Which claims appear, in what order, what gets left out, where the emphasis falls — all of that was settled by something that does not know what you think. You can rewrite every sentence and the essay remains an essay you did not write, because the essay is the argument, not the sentences.

If you draft and the model attacks, you keep the argument and get something most students never have: a reader who responds immediately, at two in the morning, and is not tired of you.

Same tool. Same finished polish. Completely different piece of work.

A working sequence

1. Think out loud — talk to it. This stage is fine and genuinely useful. But make it the one asking:

```
I'm writing about why Bolivia's water privatisation in
Cochabamba collapsed. Don't tell me your view. Ask me
questions until I've said something specific enough to
argue.
```

2. Outline — yours, always. One sentence per paragraph, and the sentences must be claims, not topics. "Colonial land registration" is a topic. "Colonial land registration created the ownership ambiguity that the 1998 reform then failed to resolve" is a claim. Ten claim-sentences in order *is* your essay, and it is the part that carries the mark.

3. Draft — yours, and badly. Fast, ugly, out of order. The bad draft is not a lesser version of the good one; it is the mechanism by which you discover that your third point actually undercuts your second. That discovery cannot happen to a reader.

4. Critique — theirs, and hard. Now the tool earns its place:

```
What is my actual thesis, in one sentence, based only on
what I wrote? Not what I seem to have intended.
```

```
Which paragraph would a hostile examiner attack first, and
what would they say?
```

```
List every claim I make that I haven't evidenced.
Don't fix them.
```

```
Where does the essay stop supporting the thesis and start
just being about the topic?
```

That first prompt is the most useful in this course. The gap between the thesis you meant and the thesis you wrote is where most marks are lost, and you cannot see it yourself because you can see your intention.

5. Line edit — mixed, carefully. Fine to get grammar and clarity help. Watch for flattening: models smooth prose toward a comfortable average, and the average is dull.

6. Read it aloud. Yourself. Any sentence you would not say is a sentence you did not write.

What machine prose sounds like

Learn the tells, so you notice when your own writing drifts toward them.

Groups of three. "Not only X, but Y." Words like *delve*, *tapestry*, *multifaceted*, *underscore*. "It is important to note that." Paragraphs that end by summarising themselves. Hedging in both directions so nothing is claimed. Sentences of uniform medium length, one after another, forever.

Your voice lives mostly in three places: **rhythm** (short sentence after a long one), **specificity** (the actual number, the actual example — a shopkeeper in Kano, ₦40,000, a Tuesday), and **willingness to say a thing flatly** without softening it into safety.

A model asked to preserve your voice will preserve your vocabulary and quietly regularise the rhythm. Rhythm is where it goes.

If English is not your first language

This is where the tool is most valuable and the advice is most often useless. You are not trying to keep a stylistic flourish. You are trying to be judged on your argument rather than your prepositions. That is fair, and refusing help is not noble.

The distinction that matters: **ask for corrections with reasons, not rewrites.**

```
Correct only grammar and word choice. Keep my sentence
structure. List each change and why, so I learn it.
```

The list is the point. It is a personal error catalogue, and after a term of it, the same corrections stop appearing. A rewrite gives you a cleaner essay this week and nothing at all next week.

And there is a second reason to insist on your own structure. Detection systems — the subject of the next lesson — flag simple, regular, low-variance prose as machine-written, and they flag non-native writers far more often than native ones. Sounding more like yourself is, awkwardly, also the safer position.

BEFORE YOU MOVE ON

Two essays arrive with equal polish. Amara wrote her draft and asked for a critique. Tom asked for a draft and then rewrote every sentence in his own words. What has Tom lost that the rewriting cannot recover?

A. Accuracy, since he cannot verify claims he did not research himself.

B. The argument itself — which claims appear, in what order, and what is omitted was decided before he wrote a word.

C. Nothing that matters. Voice is word choice and sentence rhythm, and he replaced all of it.

D. Nothing academically; his only exposure is a procedural one about disclosing the tool.

48 · 8 min

Where the thinking happens

THE ONE THING TO KEEP

Ideas are finished by being ordered, so assistance at the claim and structure stages removes the thinking itself — the outline is the essay, and everything downstream can be helped safely.

Writing is not transcription

There is a common model of writing in which you have the ideas first and then write them down. It is wrong, and it matters here more than usual.

What actually happens is that ideas are half-formed until you try to put them in order, and the attempt is what finishes them. You discover the argument does not work in the third paragraph. You find that two points you thought were separate are the same point. You realise the thing you assumed was background is the actual claim. None of that was available before you wrote; it happened *in* the writing.

This is why *I know what I want to say, the model just puts it into words* is a mistaken description of what has occurred. If a model produced the words, it did the ordering, and the ordering is where the discoveries live.

The stages, and what each one costs

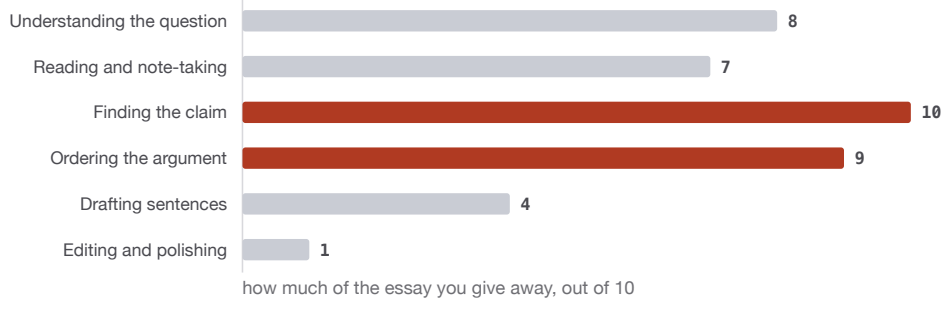
Writing an essay has roughly six stages. They are not equally valuable, and knowing which is which decides everything about where help is safe.

1. **Understanding the question.** Yours. Getting this wrong invalidates everything downstream, and it is the most common cause of a bad mark.
2. **Reading and note-taking.** Yours, per Module 5.
3. **Finding the claim.** Yours, absolutely. This is the essay.
4. **Ordering the argument.** Yours. This is where the thinking finishes.

5. **Drafting sentences.** Mixed. The argument is fixed by now; this is execution.
6. **Editing and polishing.** Safest place for help, by a wide margin.

The gradient runs one way. Assistance at stage six costs you almost nothing. Assistance at stage three costs you the essay, even if every word is later rewritten in your own style — because the piece is now a defence of somebody else's claim.

What help costs you, stage by stage



Not a measurement: the gradient this lesson describes, put on one scale so the shape is visible. Help at the bottom costs almost nothing and is what a writing centre has always done. Help at finding the claim and ordering the argument costs the essay, even if every word is later rewritten in your own style, because the piece is now a defence of somebody else's claim.

The outline is the essay

If your outline is genuinely yours — the claim, the order of moves, the evidence chosen for each — then the piece is yours, and everything downstream can be assisted without dishonesty or loss.

Which makes the practical test simple. Look at your outline. Can you say, for each move, why it is there and why it comes at that point? If yes, proceed. If it came from a model and you are reconstructing the reasons afterwards, you are editing an argument you did not make, and you will feel the difference in the third paragraph when you cannot defend a transition.

Writing to find out what you think

A practical technique that predates all of this and works better than anything else when you are stuck. Write badly, on purpose, fast, with no structure. Two

hundred words on what you think the answer is, ungrammatical, contradicting itself.

Then read it. Somewhere in there is a sentence that is more definite than the rest. That is usually your claim, and you did not have it ten minutes ago.

This works because it lowers the cost of a bad sentence to nothing. Most writing paralysis is the requirement that the first sentence be good. A model solves that by supplying a good first sentence, which removes the paralysis and the discovery together. Free-writing solves it by removing the requirement, which keeps both.

Talking as a route in

If you think better out loud, dictate. Every phone transcribes free, and many assistants take voice. Talk for five minutes about what you think the answer is, get the transcript, and edit it into an outline yourself.

This is a good use because the content is yours. The machine transcribed; it did not compose. The distinction to hold: it is doing the typing, not the thinking.

The specific loss

When students hand over stages three and four, the essays that come back share a signature. They are competent, well-organised, and say nothing. They land in the middle band and stay there. Markers describe them the same way every time: *no argument*, *descriptive*, *does not engage with the question*.

That is not a detection problem. It is the accurate description of a piece whose claim was the average of what has been written on the topic, because that is what the process produces. And it is why the students who use these tools well are conspicuously careful about which stage they use them at.

BEFORE YOU MOVE ON

A student develops the argument, writes the outline, then asks a model to draft the prose from it and rewrites every sentence in their own style. What has been lost?

A. The voice, since rewriting a generated draft sentence by sentence cannot remove its underlying rhythm.
B. The process record, since a generated draft cannot demonstrate authorship in a version history.
C. Very little, because the claim and the ordering — the stages where the thinking happens — were done first.
D. The discoveries that happen during drafting, since some of the argument is always finished at the sentence level rather than in the outline.

49 · 8 min

The blank page, without outsourcing it

THE ONE THING TO KEEP

Blank-page paralysis has four distinct causes with four different fixes, and the request that removes the essay is the short one — a suggested thesis is fifteen words and it is the entire piece.

Why the first sentence is hard

Staring at a blank document is not laziness and it is not lack of ideas. It is usually one of four things, and each has a different fix. Diagnosing which one you have takes thirty seconds and saves an evening.

You do not know enough yet. The honest and most common cause. Nothing will make writing possible; go and read three more things and take notes. If you cannot say what the argument is in one sentence out loud, you are not ready to write, and no technique substitutes for that.

You know enough but have not decided what you think. Free-writing fixes this. Two hundred words of deliberate rubbish, then find the definite sentence.

You have decided but the first sentence has to be perfect. The standard is the problem. Write a placeholder — literally *THIS ESSAY ARGUES THAT* in capitals — and carry on. Come back at the end when you know what the essay actually did.

You are avoiding it because it is unpleasant. Not a writing problem. Twenty-five minutes on a timer with the internet off.

Fixes that do not cost you the essay

- **Start in the middle.** There is no rule that says the introduction comes first. Write the paragraph you are most sure of, and the surrounding structure often becomes obvious.
- **Write the topic sentences only.** One sentence per paragraph, stating what that paragraph argues. Six of those is an outline you can defend, and the essay is then a filling-in exercise.
- **Explain it to someone.** Out loud, to a person or a recording. Transcribe. Edit.
- **Write the conclusion first.** If you know where you are going, the argument that gets there is easier to construct.
- **Change the medium.** Paper, a different font, a different room. Trivial and it works often enough to be worth trying before anything else.

The legitimate role for a model at this stage

The safe uses are all interrogative rather than generative.

> *I have to answer this question. Here is what I think, roughly, in a hundred messy words. Ask me five questions that would force me to be more specific. Do not answer them and do not suggest a thesis.*

That is a research supervisor's move, and it works. Questions push you towards your own position; a suggested thesis replaces it.

Also safe:

- *What is this question actually asking? What are the ambiguous words in it?* Genuinely useful and often reveals that you were about to answer a different question.
- *What would somebody need to establish in order to answer this well?* A map of the work, not the work.
- *Here are my six topic sentences. Which two are making the same point, and where does the argument jump?* Diagnosis on your structure.

The unsafe one, precisely

Give me a thesis for this question and *give me an outline* are the two requests that remove the essay. They feel like small favours because the output is short. The length is misleading — a thesis is fifteen words and it is the entire piece.

If you have already asked and are now looking at a suggested thesis: do not simply use it, and do not simply discard it either. Write down two other positions somebody could take on the question, then choose between all three and write out the reason. You have converted a supplied answer back into a decision, which is what the stage was for.

When the deadline is tonight

Honest practical advice for the situation everybody is actually in sometimes. The order that saves the most:

1. Twenty minutes deciding what you think. This is not the step to cut; an essay with a clear claim and thin evidence beats a well-evidenced essay with no claim, in every marking scheme.
2. Topic sentences for every paragraph, written by you. Fifteen minutes.
3. Draft fast and badly. Do not edit while drafting.
4. Use the time you have left on the parts that carry marks: the argument's clarity and the evidence, not the introduction's elegance.
5. Check the references. All of them. This is the step people skip under pressure and the one with the worst consequences.

Notice that nothing in that list is *ask a model to write it*. Under time pressure, that route produces a mid-band mark at best and a misconduct case at worst, and it takes about the same amount of time as doing it.

BEFORE YOU MOVE ON

A student cannot start writing and cannot state their argument in one sentence out loud. What does this indicate?

A. That the first sentence standard is too high, and a placeholder will unblock them.

B. That the question is ambiguous and needs clarifying before any writing can sensibly begin, which is a comprehension problem rather than a writing one.

C. That they should free-write to discover what they think, since the argument is present but unarticulated.

D. That they have not read enough yet, and more writing technique will not help.

---

50 · 8 min

Structure that carries an argument

THE ONE THING TO KEEP

Shuffle your paragraphs — an argument breaks when reordered and a list does not — and the marks live in analysis and in stating the objection at its strongest.

The test that distinguishes an argument from a list

Take your paragraphs and shuffle them. If the essay still makes sense, it is a list of things about a topic, not an argument.

A real argument has paragraphs that depend on each other. Paragraph three establishes something paragraph four uses. Reorder them and paragraph four now refers to something that has not happened. That dependency is the differ-

ence between the middle band and the top one, in every essay subject and in most report writing.

This is also the clearest structural signature of generated prose. Generated essays are almost always reorderable, because they are assembled as a set of relevant sections rather than a chain.

Shuffle the paragraphs and see which one you have

A list about a topic

- Paragraphs of even length, one per sub-topic
- Each stands alone; none needs the one before it
- Reorder them and the piece still makes sense
- Evidence presented, then left for the reader to interpret
- Objections stated weakly enough to dismiss for free

An argument

- Paragraph three establishes what paragraph four uses
- Reorder them and paragraph four refers to something that has not happened
- Every topic sentence carries a claim, not a heading
- Analysis after every piece of evidence, saying why it counts
- The objection at its strongest, then answered or conceded

Reorderability is the clearest structural signature of generated prose, because it is assembled as a set of relevant sections rather than as a chain. It is also the difference between the middle band and the top one in most rubrics, and analysis is usually the largest single band — which is why the paragraph that presents evidence and moves on is the commonest way marks are lost.

The shape underneath

Most good academic arguments have the same skeleton:

1. **The question, made precise.** Including which interpretation of an ambiguous term you are adopting, and why.
2. **The claim.** Contestable, specific, stated early.
3. **The strongest support**, first — not last, and not buried in the middle where students traditionally put it.
4. **The best objection**, stated at its strongest.
5. **The response**, which may concede ground. Conceding a real point is a strength; it demonstrates you understood the objection.
6. **What follows**, including what your claim does not establish.

Step four is where marks are won and lost. A straw objection — *some may argue X, however this ignores Y* — costs you rather than gains you, because it demonstrates that you either did not find the real objection or avoided it.

Using a model on structure without giving it away

The safe requests are all diagnostic and all operate on your text.

> *Here are my topic sentences. Which two make the same point? Where does the argument jump without support? Which paragraph could be removed with no loss?*

> *Here is my draft. Reorder my paragraphs at random and tell me whether the essay still works. If it does, tell me which dependencies are missing.*

> *What is the strongest objection to my claim? Not a general counterargument — the specific one somebody who knows this literature would raise.*

> *I claim X. What would somebody have to believe for X to be false?*

Every one of these gives you a diagnosis and leaves the writing with you. Compare with *restructure my essay*, which hands over stage four from the previous lesson.

The paragraph, mechanically

A paragraph that works usually has four parts, and checking them is quick:

- **A claim** — the topic sentence, stating what this paragraph argues, not what it is about.
- **Evidence** — specific, cited, with a page or a figure.
- **Analysis** — why the evidence supports the claim. This is the part students omit, and it is where the marks are.
- **A link** — how this leads to the next paragraph.

The most common failure is a paragraph that presents evidence and then moves on, leaving the reader to work out why it mattered. The marker will not do that work for you, and analysis is usually the largest single band in a rubric.

A useful check on your own draft: highlight the analysis sentences. If a paragraph has none, it is description.

Signposting, and how much is too much

Academic writing needs enough signposting that a reader knows where they are, and students overcorrect in both directions. *In this essay I will argue X, then Y, then Z*, followed by *having argued X, I will now turn to Y*, is a common generated pattern and reads as padding.

Better: signpost by making the connection explicit rather than announcing the structure. *If the fiscal explanation holds, then the timing of the reforms should track revenue crises rather than elections* — that sentence both moves the argument on and tells the reader what is coming, in one move.

Length, and what to cut

When you are over the word limit, cut in this order:

1. Anything that summarises what you are about to say or have just said.
2. Background the reader of your course already has.
3. Quotations longer than a sentence, which are almost always cuttable to a phrase.
4. Any paragraph whose removal you cannot immediately object to.

What you never cut: analysis, the objection, and the specificity of your evidence. Students under a limit tend to cut exactly those, because they are the hard parts and look expendable, and the essay drops a band.

BEFORE YOU MOVE ON

What does it indicate if an essay still reads perfectly well after its body paragraphs are shuffled?

A. That it is a list of points about a topic rather than an argument with dependencies.
B. That the signposting is doing its job, since the reader can follow the essay from any starting point.
C. That the paragraphs are well constructed, since each stands on its own with a claim, evidence and analysis.
D. That the essay is descriptive rather than analytical, because analysis by its nature produces material that could appear in any order.

51 · 8 min

Editing your own prose

THE ONE THING TO KEEP

Accept only changes you can explain, ask for diagnosis rather than replacement, and reject the standard drift towards nominalisation, hedging and abstraction that makes edited prose lifeless.

The safest place for help, and the rules for it

Editing prose you wrote is the one stage where assistance is nearly free of cost, because the argument is already fixed. It is also where a lot of students accidentally lose their voice, by accepting everything offered.

The governing rule: **accept changes you can explain.** If you cannot say why the new version is better, do not take it. That single rule keeps the prose yours and turns each edit into a small lesson in writing.

Ask for diagnosis, not replacement

> *Here is my paragraph. Do not rewrite it. Tell me: which sentence is unclear and why, where I have used two words for one idea, which sentence is longest and whether it needs to be, and where I have asserted something I have not supported.*

Diagnosis, then you fix it. This is slower and it is the version where you get better at writing. A rewritten paragraph handed back teaches you nothing, because you cannot see the reasoning that produced it.

When you do want a rewrite for comparison, ask for it *alongside* yours with the changes named:

> *Rewrite this one sentence, then list every change you made and why.*

Then decide, change by change. Usually two of the six are improvements and the rest are the model's preferences, which are not yours.

What the changes usually are

Assistants edit towards a particular register: more formal, more hedged, more abstract, more evenly-weighted sentences. Some of that helps academic writing and some of it is exactly what makes prose lifeless.

Watch for these specifically, and reject them by default:

- **Nominalisation.** *We investigated* becoming *an investigation was undertaken*. Longer, vaguer, and it hides who did what.
- **Hedge inflation.** *This suggests* becoming *this may potentially suggest that in some cases*. Three hedges where one was honest.
- **Elegant variation.** Calling the same thing by three different names to avoid repetition. In academic writing, repeat the term; the reader needs to know it is the same thing.
- **Sentence-length flattening.** Everything becoming the same medium length. Rhythm comes from variation — a long sentence doing complicated work, then a short one landing it.

- **Abstraction creep.** Specific examples replaced by general statements. This is the most damaging one, because specificity is what earns marks.

The edits worth making yourself

The list that improves almost any student essay, in order of return:

1. **Cut the first sentence of every paragraph if it is throat-clearing.** *It is important to consider that…* Start at the claim.
2. **Find every *it is* and *there are*.** Most can be replaced with a real subject and verb. *There are several factors which influence* becomes *Three factors influence*.
3. **Check every *this* has a noun after it.** *This shows* — this what? Ambiguous *this* is the most common source of unclear academic prose.
4. **One idea per sentence** where the idea is difficult.
5. **Read it aloud.** You will hear what you cannot see. This is the single most effective editing technique and it costs nothing.

Grammar and spelling tools

Spellcheck, LanguageTool (free, open source, and better than most paid options for non-English languages), and the grammar checkers built into word processors are ordinary writing tools and almost universally permitted. Grammarly's free tier is similar; its paid rewriting features are generative and sit in a different category, and some institutions treat them differently. If your policy distinguishes, the line is usually between correcting an error you made and producing text you did not write.

For students writing in a second language, this category of tool is genuinely levelling and you should use it without guilt. Keep your drafts, for the reasons in Module 7.

The read-aloud test, one more time

If a sentence is hard to say, it is hard to read. If you run out of breath, it is too long. If you stumble on a phrase, so will your marker.

Every word processor now has a read-aloud function, which is free and catches things your eye skips because you know what it was supposed to say. Ten minutes of listening to your own essay before submitting will catch more than any amount of re-reading.

BEFORE YOU MOVE ON

An assistant changes "We measured reaction times in 40 participants" to "Reaction time data were collected from a participant sample." Why should this be rejected?

A. Because the passive voice is not acceptable in academic writing, which requires active constructions throughout.

B. Because it nominalises the action and drops the specific number, losing precision.

C. Because the original was shorter, and concision is the primary criterion in academic prose.

D. Because it removes the first-person plural, which is now preferred in most scientific style guides as clearer about who did what.

---

52 · 8 min

Your voice, measurably

THE ONE THING TO KEEP

Voice is the visible record of a person making choices — which example, which objection, which concession — and it is what markers reward, so build a corpus of your own writing to compare drafts against.

Voice is not decoration

Students hear *keep your voice* as advice about personality, which makes it sound optional in academic writing. It is not. Voice in an essay is the set of

choices that shows a particular person thought about this: which example you reached for, what you found worth objecting to, where you conceded, what you thought was obvious enough to skip.

A marker reading two hundred essays on the same question is looking for exactly that. It is what distinguishes a first from a two-one far more reliably than vocabulary does.

Build a corpus of your own writing

A concrete exercise, and unusually useful. Collect four or five pieces of writing that are unambiguously yours — old essays, long messages, anything written before you used these tools. Read them together and note:

- **Sentence length.** Do you run long and subordinate, or short and declarative?
- **Where your examples come from.** Do you reach for history, for numbers, for the everyday?
- **Your standard moves.** Do you open with a concession? A definition? A specific case?
- **Words you actually use.** Everybody has a small set of characteristic connectives and qualifiers.
- **What you find funny or objectionable.** It shows even in formal writing.

Now you have something to compare a draft against, which converts a vague worry into a check you can perform.

A reasonable use of a model, on your own material:

> *Here are four pieces I wrote. Describe my writing: sentence structure, characteristic constructions, register, how I use examples. Be specific and do not flatter.*

Students are frequently surprised. Most people have never had their own prose described back to them.

The tells of generated prose

These are patterns, not proof — this course is clear elsewhere that nobody can reliably detect authorship from text. But as things to remove from your own writing, they are useful:

- **Tricolon everywhere.** Three-item lists in every sentence, regardless of whether there are three things.
- **Uniform paragraph length.** Real writing has a paragraph that is two sentences because that was all it needed.
- **The balanced summary ending.** *Ultimately, while X offers benefits, careful consideration of Y remains essential.* Says nothing.
- **Hedged everything.** No sentence commits.
- **Abstraction over instance.** Claims about *various factors* and *a range of approaches* where a specific example belonged.
- **Preamble.** A first paragraph restating the question before beginning.
- **No irregularity.** Nothing surprising, no aside, no sentence that took a risk.

The common thread is the absence of a person making choices. That is the thing to add back, and it is also the thing that earns marks.

Register is not vocabulary

A persistent misconception: that academic writing means long words. It does not. Academic writing means precise, and precision often means the plain word.

What academic register actually requires: hedging claims to the strength of the evidence, defining terms you are using in a specific sense, attributing positions to whoever holds them, and being explicit about the logical relationship between sentences. None of that requires *utilise* or *heretofore*.

If a model raises your register, check whether it raised your precision. Usually it did not, and inflated vocabulary is easier for a marker to see through than plain prose is.

Reading is how voice develops

The uncomfortable long answer. Your prose is built from what you have read, and if you mostly read generated summaries, that is what your writing will sound like.

So read real writing in your field: actual papers by people known for writing well, and good non-fiction generally. Notice a sentence you like and work out why. That is the whole method, it is slow, and there is no substitute.

Keeping it while accepting help

The practical arrangement, drawn from everything above:

- Write the first version yourself, always. It can be terrible.
- Ask for diagnosis rather than rewriting.
- Accept only changes you can explain.
- After a heavy editing session, read the piece aloud. If it does not sound like you, you accepted too much. Go back to your draft and take fewer changes.

That last check takes two minutes and is the one that keeps a term of assisted editing from slowly converting your prose into everybody's.

BEFORE YOU MOVE ON

Why is inflated vocabulary a worse strategy than plain prose in academic writing?

A. Because markers penalise unnecessary complexity directly under a clarity criterion in most rubrics.

B. Because subject-specific terminology should only be used once it has been formally defined in the text, and inflated vocabulary usually appears undefined.

C. Because complex vocabulary is a known signature of generated text, so it invites suspicion of AI use.

D. Because register in academic writing is about precision, and unearned vocabulary is easier for a marker to see through.

53 · 8 min

Feedback that is actually useful

THE ONE THING TO KEEP

Paste the mark scheme, name a specific hostile reader, forbid rewriting and encouragement, and take feedback on structure before sentences — then use the comments on your last five assignments to find the fault that repeats.

Encouragement is not feedback

Ask a model what it thinks of your essay and you will get praise, then some gentle suggestions, then more praise. This is a consequence of how these systems are tuned, and it is worse than useless: it consumes your time and returns a false signal about quality.

The fix is to remove the conditions that produce it. Ask for something specific, adversarial and structured, and refuse the encouragement explicitly.

> *Act as a marker who has read two hundred of these and is behind schedule. Identify the three things costing the most marks, in order. Quote the sentence that demonstrates each. Do not list strengths. Do not rewrite anything. Do not be encouraging.*

That produces feedback. The framing does the work: a specific reader, a specific job, a fixed number of items, evidence required, and the failure modes named and prohibited.

Ground it in the actual criteria

The single highest-value thing to paste alongside a draft is the assessment brief and the mark scheme. Nearly nobody does this, and it transforms the output.

> *Here is the brief, the mark scheme, and my draft. For each criterion, state which band this currently sits in and quote the evidence for that judgement. Then give the three changes that would move the most marks. Do not rewrite.*

This works because it is grounded, and because it converts a vague quality question into a rubric-matching task, which is the kind of thing these systems do well on supplied text.

It is also a genuinely permitted use in nearly every institutional policy, since it produces no text for your submission. Check yours if unsure.

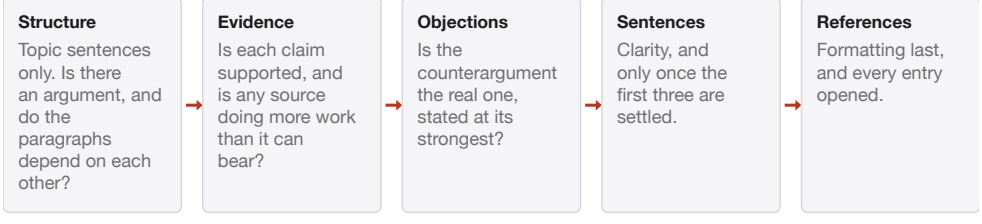
The order of feedback rounds

Getting feedback in the wrong order wastes it. Structure over sentences, always.

1. **Structure.** Topic sentences only. Is there an argument? Do the paragraphs depend on each other?
2. **Evidence.** Is each claim supported? Is any source doing more work than it can bear?
3. **Objections.** Is the counterargument the real one, at its strongest?
4. **Sentences.** Clarity, only after the first three are settled.
5. **Formatting and references.**

Polishing sentences in a paragraph you will later delete is the most common way students waste an evening. Do not edit prose until the structure is fixed.

The order of feedback rounds



Polishing sentences in a paragraph you will later delete is the commonest way to waste an evening, so do not edit prose until the structure is settled. Paste the brief and the mark scheme alongside the draft and the whole exercise changes character: it becomes a rubric-matching task on supplied text, which these systems do well, and it is permitted under nearly every policy because it produces no text you submit.

Human feedback is still better, and there is more of it than you think

- **Office hours.** Chronically underused. Ten minutes with the person marking your work is worth more than any amount of AI critique, because they know what they are looking for and they will tell you.
- **Ask a specific question.** Not *is this any good* but *is my third paragraph doing the work I think it is*. Specific questions get specific answers and take less of everyone's time.
- **Writing centres.** Most universities have one, free, staffed by people whose job is exactly this.
- **A peer.** Swap essays with someone on a different question. A reader who does not know the topic will find every place your reasoning was implicit, which is precisely what a marker will find.

Feedback on returned work

The feedback you already have and did not use: your marked assignments. Most students read the grade and close the file.

A better routine. Collect the comments from the last five pieces of work into one document, and look for what repeats. Almost everybody has two or three persistent faults — *not enough analysis*, *unsupported claims*, *doesn't answer the question* — and fixing one of them raises every future mark.

A model is genuinely useful here: paste the comments and ask what patterns appear across them. That is analysis of supplied text, which is the safe category, and it surfaces things you have stopped seeing.

Feedback you give to others

Reading somebody else's draft is one of the fastest ways to improve your own writing, because faults are far easier to see in another person's work than in your own. You will notice the unsupported claim and the missing analysis instantly, and then find yourself noticing them in your own draft the following week.

Give it the way you would want it: name the two things costing the most, quote the sentence, and say what would fix each. Do not rewrite their prose.

The one to be wary of

Asking a model to predict your grade. It will give a number, it will sound confident, and it is unreliable: it does not know your institution's standards, your cohort, or your marker. Students act on these numbers, decide a draft is fine, and submit.

Use it for *which criteria are weakest and why*, which is diagnostic and checkable against the rubric. Ignore the number.

BEFORE YOU MOVE ON

Why should feedback on structure come before feedback on sentences?

A. Because sentence-level feedback is less reliable, since models edit towards their own register rather than yours.

B. Because structural problems are harder to fix, so they need the most remaining time before a deadline.

C. Because polishing prose in paragraphs you will later delete wastes the effort entirely.

D. Because a marker forms their judgement of an essay from its structure first, so structural weaknesses cost more marks than unclear sentences do.

54 · 8 min

Presentations, posters and visuals

THE ONE THING TO KEEP

Write the talk first and let slides follow from it, and never generate an image that carries information — a generated diagram is a picture of what a diagram looks like, with plausible wrong labels.

The deck is not the talk

The most common failure in student presentations has nothing to do with AI: slides full of text, read aloud. A model will produce that instantly and in volume, so the failure is now available faster.

What is assessed in a presentation is almost never the slides. It is whether you can explain something, in order, to people, and answer questions about it. The slides support that. If your slides could be read without you, you have written a document and are now reading it aloud, which is the least useful form either could take.

Making slides that work

- **One idea per slide.** If it needs two, it needs two slides.
- **The heading is the point, not the topic.** Not *Results*, but *Response rate doubled in the treatment group*. A reader skimming your headings should get the argument.
- **Six words on a slide is plenty.** The words in your mouth are the content.
- **A figure beats a bullet list**, if the figure shows something.
- **Anything you would read aloud does not go on the slide.**

A legitimate and useful AI request:

> *Here is my talk, written out. Suggest where a slide should change and what single point each slide should carry. Do not write slide text.*

The talk exists first. The slides follow from it. Working the other way — slides first, then working out what to say — is how you end up reading.

Free tools, named against the paid ones

- Slides: PowerPoint and Keynote → **LibreOffice Impress**, **Google Slides**, or **reveal.js** if you want them in a browser and in version control.
- Diagrams: Visio, Lucidchart → **draw.io** (free, browser or desktop), **Inkscape** for anything that needs to be publication quality, **Mermaid** for diagrams written as text.
- Charts: Excel → **LibreOffice Calc**, Google Sheets, or **matplotlib** in Python for anything you will need to regenerate.
- Posters: InDesign, Illustrator → **Inkscape** or **Scribus**, both free and both entirely capable of an A0 conference poster.
- Images: Photoshop → **Photopea** in a browser, **GIMP**, **Krita**.
- Video: Premiere → **DaVinci Resolve** free edition, **Shotcut**, **Kdenlive**.
- Audio: → **Audacity**.

Every one of these runs on a modest machine, and several run in a phone browser.

Generated images in academic work

A distinction worth making clearly.

Decorative images — a background, an illustration on a title slide — are usually fine, and you should say they were generated if your institution asks.

Images that carry information — a diagram of a mechanism, a chart, a figure — must not be generated. A generated diagram is a picture of what a diagram looks like, and it will contain plausible wrong labels, arrows in the wrong direction, and structures that do not exist. Draw it in Inkscape or draw.io, or plot it from data.

Anything representing your data must come from your data. Generating or beautifying a figure that represents results is research misconduct, not a style choice, and the line is not blurry.

And copyright here is genuinely unsettled. Whether generated images are protected, who owns them, and whether their production infringed the works they were trained on are live legal questions with different answers in different countries and pending litigation in several. If you are producing something that will be published rather than submitted, that uncertainty is a reason for caution. This is not legal advice, and anyone who tells you the position is settled is ahead of the courts.

Rehearsal, which is the actual preparation

Slides get all the attention and rehearsal decides the outcome.

- Say it out loud, standing, at least three times. Silently reading through is not rehearsal.
- Time it. Almost everyone overruns, and being cut off mid-argument is the worst possible ending.
- Rehearse the transitions between slides specifically. That is where people stumble.
- **Prepare for questions**, which is where a model is genuinely excellent: paste your talk and ask for the ten hardest questions an examiner would ask. Then answer them out loud. This is the single best use of these tools for a presentation, and it is exactly the viva preparation from Module 8.

Posters

Same principles, one addition: a poster is read standing up, at a distance, by somebody deciding in four seconds whether to stop. Title readable from three metres, one clear finding, and the rest is a conversation starter rather than a paper in small type. Scribus and Inkscape do this free, and your department's plotter usually prints for nothing or close to it.

BEFORE YOU MOVE ON

Why is a generated diagram of a biological mechanism unacceptable in a presentation, while a generated decorative background is usually fine?

A. Because the diagram carries information, and generation produces plausible labels and arrows rather than correct ones.
B. Because the diagram is more likely to breach copyright, since mechanisms are drawn from published textbook figures.
C. Because diagrams are assessed under the content criteria while decoration is not, so only the diagram can cost marks.
D. Because a scientific figure must be reproducible from a stated source or dataset, and a generated one has no provenance that a reader could follow up.

55 · 7 min

Group work, and shared rules

THE ONE THING TO KEEP

Agree the group's AI rule in writing before the first session, work in a shared document with version history, and require each person to confirm they opened every source in their own section.

The submission is joint and the consequences are not

When four people submit one document, all four are answerable for all of it. If one section contains a fabricated citation or generated prose that the group's declaration did not cover, the investigation involves everybody, and *I did not know* is a weak position that also damages a friendship.

This is the most common way careful students end up in integrity meetings, and it is almost entirely preventable by a fifteen-minute conversation nobody wants to have at the start.

The conversation, in one message

Send it before the first working session:

> *Before we start: how are we using AI on this? My suggestion — fine for generating questions, critique of our own drafts, translation, and checking understanding. Not for prose that goes in the document. Everyone opens every source they cite. Everyone writes their own section. If anyone wants to work differently, say now rather than at the end.*

Awkward for about ninety seconds. Then it is settled, in writing, with a timestamp, and everybody knows the standard.

What matters is not that everybody adopts the strictest possible rule. It is that the rule is explicit and shared. A group where two people think generated drafting is fine and two think it is cheating will have a problem, and it will arrive at the worst moment.

Check the actual policy first

Course policies on group work sometimes differ from individual-work policies, and some require a group declaration listing what each member did and what tools were used. Read yours before agreeing anything, so the group rule is at least as strict as the institutional one. If your policy is silent on group work, ask the module leader in writing — the reply is useful evidence later.

Working in a shared document

Practical arrangements that prevent most problems:

- **Google Docs or Word online**, with version history on. Every contribution is attributed and timestamped automatically. This is the single most useful thing you can do, and it costs nothing.
- **Never paste a finished block into a shared document.** Write in it, or paste in pieces with edits, so the history shows work rather than an arrival. A large block appearing at once is what triggers questions, whoever wrote it.
- **One person owns each section**, named at the top of the draft.

- **One person does a consistency pass at the end**, and says so.

The reference list is where it goes wrong

Combining four bibliographies is the moment fabricated citations enter a document, because the person merging did not read anybody else's sources and assumes they were checked.

The rule that fixes it: **whoever cites it, opens it, and says so.** Before submission, each person confirms in the group chat that every reference in their section was opened and that they can say which page they used. That message is your record, and it takes two minutes.

If someone will not confirm, the reference comes out. This is not distrust; it is exactly the standard everybody would want applied to their own name on the cover.

Dividing the work, revisited

From the study-groups lesson, but with the assessment consequence attached. Splitting topics so each person only learns a quarter is efficient for producing a document and poor for learning, and in any group assessment with an individual component — a viva, a reflective piece, a question in an exam — it is directly costly.

The arrangement that survives: split the *writing*, share the *understanding*. Everybody reads the whole draft, and everybody should be able to defend any section. Half an hour before submission, take turns explaining each other's sections. It is the fastest possible check for anything nobody actually understands, which is also the fastest possible check for anything nobody actually wrote.

If it goes wrong

If you find, late, that a section contains work you are not comfortable submitting: raise it with the group first, immediately, in writing. If it cannot be resolved, tell the module leader before submission rather than after. Disclosing

early is treated completely differently from being found later, at every institution, and it is also the only version you will feel fine about afterwards.

BEFORE YOU MOVE ON

Why is "whoever cites it, opens it, and confirms so before submission" the specific rule that protects a group?

A. Because it distributes the checking workload evenly, which makes verification feasible for a long reference list.

B. Because institutional policies on group work generally require an individual declaration from each member about the sources they personally consulted.

C. Because it creates a written record that identifies who is responsible if a fabricated reference is found.

D. Because merged bibliographies are where unchecked references enter, since the person combining them has not read anyone else's sources.

CASE STUDY · MODULE 6

The section nobody would claim

A four-person BBA marketing project in Indore, two days before submission, with one section that does not read like the person who wrote it.

The report was six thousand words on competitor positioning in the regional snack food market, split four ways, and due on Friday. Deepak's fourteen hundred words arrived at eleven on Wednesday night, in one message, complete and formatted.

Nikita Shah, who was co-ordinating, read it twice. Nothing in it was wrong, which was part of the difficulty. Every paragraph was about the same length. Every claim was hedged in both directions, so that nothing was actually asserted. Several paragraphs ended by summarising themselves. The section closed with a sentence about careful consideration of multiple factors going forward, and the phrase multifaceted appeared twice.

She did the only test she knew that does not depend on a feeling about style. She copied the paragraphs into a blank document and put them in a different order. The section still read perfectly well. That told her what the register had already suggested: it was a set of relevant sections rather than an argument, because nothing in paragraph four depended on paragraph three having happened first.

Then she checked the references, because that was quicker than arguing about prose. Eleven citations. Ten resolved through the DOI resolver or the library catalogue. One was a 2021 article in a marketing journal, with real authors who work in that field, a plausible title, a real journal, and a DOI that returned an error.

The surrounding facts mattered, and they pulled in different directions. Deepak had been genuinely ill for most of the preceding fortnight and had told nobody. He had also delivered on time, which nobody else in the group had managed. The group had never had the conversation about how they were using these tools: Nikita had thought about proposing it in week one and decided against because it would sound as though she distrusted people she had

known for two years. And the group was split without knowing it. She and Pooja thought generated drafting was obviously cheating. Arjun had said more than once, without anybody objecting, that everybody does it and it is fine as long as you edit it afterwards.

Four options were open at eleven on Wednesday night, and none of them was cheap.

Submit it. One document, four names, and a module brief that requires a group declaration stating what each member contributed and which tools were used. If a question is ever raised, all four are in the process together, and I did not know is a weak position that also ends a friendship.

Rewrite it herself overnight. This is the option that feels kind, and it is the one she wanted to take. It also means the declaration she signs becomes false by her own hand rather than by Deepak's, and it leaves him facing the module's individual viva, worth ten marks, on a section he did not write in either of its versions.

Ask Deepak to redo it himself in two days, while ill, knowing that the honest version would be shorter and thinner and would probably cost the group marks.

Or tell the module leader that evening, before submitting anything.

STOP HERE

Before turning the page: what would you have done, and what would it have cost if you were wrong? Write it down. The point of the next section is to compare.

WHAT HAPPENED

Nikita phoned Deepak rather than messaging. He said he had written it with help and had genuinely not thought of that as generating it, which she believed. He had not checked the DOI; he had assumed it came from somewhere.

They took the fourth option in a modified form. They emailed the module leader that evening, before the deadline, saying that one section needed rework, asking for a short extension, and saying why in three sentences. They were given seventy-two hours, with a note that the request had been recorded.

Deepak rewrote the section himself from the sources he had actually read, which turned out to be three of the eleven. It came back at nine hundred words. It was worse written and considerably better: it made one contestable claim, it had a number in it that came from a report he had opened, and two of its paragraphs could not have been swapped. The declaration listed tool use per person per section, and Nikita's own entry included having asked for structural feedback on her draft.

The report scored sixty-four, below what they had hoped, and the module leader's comment said the competitor section was thin. No misconduct process was opened. All four answered their viva questions.

Deepak did not speak to Nikita for the rest of the term. She says she would do it the same way again and that it still cost her a friend. She also says the fifteen-minute conversation in week one, which she avoided because it would have sounded distrustful, would have cost about ninety seconds of awkwardness.

Nikita reordered the paragraphs before she looked at anything else. What does that test detect, and why does it detect it?

It distinguishes an argument from a list of things about a topic. A real argument has dependencies: paragraph three establishes something paragraph four uses, so reordering breaks it, and paragraph four now refers to something that has not happened. A list survives reordering because each part stands alone. Generated prose is almost always reorderable, because it is assembled as a set of relevant sections rather than as a chain, and the same is true of a human draft that never found a claim. The test is worth running on your own work for that second reason: it does not tell you who wrote something, and it does tell you whether the thing in front of you is arguing, which is where the marks are.

Rewriting the section herself was the option that felt kindest. Make the case that it was the worst of the four.

It fails on three counts at once. The declaration the brief requires would then be false in a way she authored knowingly, which is worse than an omission and converts a problem she inherited into one she owns. It leaves Deepak facing an individual viva on a section he did not write in either version, which is the moment the whole thing surfaces with no good answer available. And it conceals the disagreement in the group rather than resolving it, so the same situation recurs on the next assessment with the same four people. The option they took has the property none of the others has: disclosing before submission is treated completely differently from being found afterwards at every institution, and it was still available at eleven on Wednesday night.

Arjun thought generated drafting was fine if you edit it. Could a group rule permit that, and what would have to be true?

It could, but only downstream of two things the group did not have. The institutional rule sets the floor: a group rule cannot be looser than the module's policy, so the first step is reading the brief and the handbook and, if they are silent, asking in writing. Second, even where a policy permits assisted drafting with disclosure, the piece has to remain the group's argument, which means the claim, the order of moves and the evidence are theirs, and every source is opened by whoever cites it. Deepak's section failed that test independently of any rule, because it had no contestable claim and one of its eleven references did not exist. What sank the group was not that Arjun's position is indefensible. It is that four people held two incompatible positions for eleven weeks without discovering it.

MODULE 6 · TEST

Check what stuck

6 questions, one right answer each. This one is for you, not for a record: answers and the reason behind each are at the back. If two questions from the same module go wrong, re-read that module before the exam rather than after.

1 Two students submit equally polished essays. What decides whether the work is theirs?

A. Whether the finished prose reads in their own register
B. Whether they or the model produced the first draft
C. How many of the model's suggested edits they accepted
D. Whether they disclosed the assistance in a statement

2 Assistance at which stage of writing an essay costs you the most?

A. Correcting grammar and spelling in prose you wrote
B. Checking a finished draft against the mark scheme
C. Rewriting a sentence whose fault you can already name
D. Finding the claim and ordering the argument

3 You shuffle your paragraphs into a random order and the essay still makes sense. What has that shown?

A. It is a list about a topic, not an argument
B. The paragraphs are well balanced in length
C. The transitions have been written too heavily
D. The evidence is spread evenly through the piece

4 What is the governing rule when a model suggests changes to prose you wrote?

A. Accept every change that shortens the sentence
B. Accept changes that raise the register of the prose
C. Accept only changes you can explain
D. Accept the rewrite and then restore your own vocabulary

5 What is the single highest-value thing to paste alongside a draft when asking for feedback?

A. Three essays by classmates who scored in the top band
B. The full reading list from the module handbook
C. A prediction of the grade the draft would receive
D. The assessment brief and the mark scheme

6 Where is the line on generated images in academic work?

A. Any generated image must be declared but may carry data
B. An image that carries information must not be generated
C. Generated diagrams are fine if the caption names the source
D. Decorative generated images are prohibited in all coursework

MODULE 7

Rules, honesty and the record

What your institution's policy actually says, how to disclose, why detectors are not evidence, what to do if you are accused, and the process record that protects you before anybody asks.

> BY THE END OF THIS MODULE YOU CAN
>
> Read your institution's AI policy and place a given piece of work on the permitted, disclosed or prohibited scale, write a disclosure statement, explain from its mechanism why a detector score is not evidence, and keep a process record that would survive an academic-integrity hearing

56 · 9 min

The Rules, And Why Detectors Do Not Work

> THE ONE THING TO KEEP
>
> Detectors measure how typical your writing is, not who produced it; keep a process record instead.

What your institution's policy probably says

Almost every policy lands in one of four postures:

1. **Prohibited for assessed work.** Any AI use is misconduct.
2. **Permitted with disclosure.** Use it, declare what you used it for.
3. **Permitted at some stages only.** Commonly: fine for research, brainstorming and proofreading; not for drafting, analysis or generating

content.

4. **Left to the module leader.** Increasingly the most common, and the most confusing.

If yours is the fourth — and there is a good chance it is — then **the university's public policy page is not your answer.** The answer is in the assignment brief, and if the brief is silent, the answer is in an email you have not sent yet.

Send it. "Is it acceptable to use an AI tool to generate practice questions and to check my grammar for this essay? I want to be sure before I start." Ask before, not after. Keep the reply. A dated email from the person marking your work is worth more than any argument you can construct later.

Also know: policies are being rewritten constantly. A rule you learned in first year may not be the rule now. Check each term.

The detectors

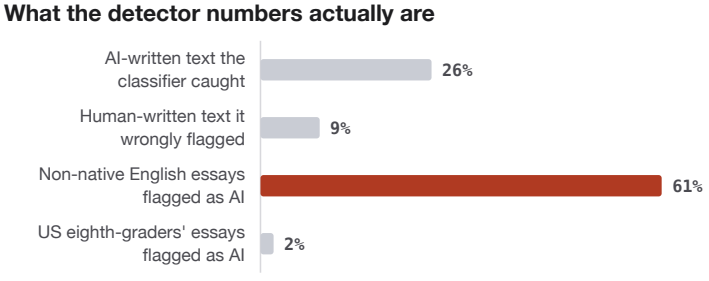
Here is the part nobody tells students clearly.

OpenAI released an AI Text Classifier in January 2023 — a detector for its own model's output. It withdrew it that July, citing low accuracy. Its own published figures: it identified about **26% of AI-written text**, and wrongly flagged about **9% of human-written text** as AI. The company that built the generator could not reliably detect the generator.

A 2023 Stanford study (Liang and colleagues, published in *Patterns*) ran seven commercial detectors over TOEFL essays written by non-native English speakers, and over essays by US eighth-graders. The eighth-graders were classified correctly almost every time. **More than half of the non-native speakers' essays were flagged as AI-generated.** The reason is mechanical: the detectors keyed on low lexical variety and simple sentence construction. A careful second-language writer produces exactly that.

Several universities responded by turning the tools off. Vanderbilt disabled Turnitin's AI detector in 2023 and said so publicly. Turnitin has claimed roughly a 1% false positive rate at document level — and even taken at face

value, across twenty thousand essays in a term, one percent is two hundred students accused of something they did not do.



OpenAI's own published figures for its AI Text Classifier, withdrawn in July 2023 citing low accuracy: the company that built the generator could not reliably detect it. The essay figures are from Liang and colleagues in Patterns, 2023, averaged across seven commercial detectors. The mechanism explains the gap exactly. A careful writer in a second language uses common words and standard constructions, and that is what the instrument measures.

Why this is not a solvable engineering problem

A detector cannot examine provenance. There is no watermark in ordinary text. What it can do is measure **statistical typicality**: how predictable each word is given the ones before it. Machine text is, on average, more predictable.

So is clear, plain, careful human writing. So is writing by someone with a smaller working vocabulary in the language of assessment. So is the prose of a student who was taught to write simply, which is what good teachers teach.

Everything that makes writing easy to read makes it look machine-made. That is not a bug to be patched. It is what the measurement is.

Protect yourself before you need to

Being wrongly accused is genuinely frightening, and it is a defence built in advance or not at all.

- **Write in something with revision history.** Google Docs and Word both keep timestamped version history. A document that grew over eleven sessions across two weeks looks nothing like one pasted in whole, and the history is inspectable.
- **Keep the mess.** Your outline, your crossed-out plan, the photo of the notebook page, the draft with the abandoned second argument.

- **Keep the sources you actually read**, with your own notes on them.
- **Be ready to talk it through.** In a viva-style conversation, the person who wrote it can explain why they cut the third section. That is the evidence that persuades.

Start the habit now. The record is worthless if you begin assembling it after the email arrives.

The thing that has to be said

Detectors failing is not permission.

The argument for doing your own work is not "you will be caught," because increasingly you will not be. The argument is the one from lesson two: the exam hall exists, the professional situation exists, and both are populated by people who can tell in ten minutes of conversation whether you know your subject. The detector was never the thing that was going to find out.

BEFORE YOU MOVE ON

A detector reports your essay as "82% AI". You wrote every word yourself. What does that number actually mean?

A. It means the tool is 82% confident, so there is about a one-in-five chance you are innocent and you now have to prove it.

B. It means 82% of your sentences matched AI-generated text held in a database, so producing your sources will clear you.

C. It is a statement about how statistically typical your sentences are, not about how they were produced — and plain, clear writing scores as typical.

D. It means your writing has been unconsciously shaped by reading a lot of AI text, which is what such tools are picking up.

57 · 8 min

Reading your actual policy

THE ONE THING TO KEEP

Find the module handbook, the brief and the integrity policy, download them, and where they are silent ask in writing and keep the reply — because several policies make non-disclosure itself the offence.

The rule that applies to you is written down somewhere

Students ask *is it allowed* and answer it from what a friend said, what a lecturer implied, or what a video claimed. The rule that will be applied to you is a document, and finding it takes ten minutes once.

Look, in this order:

1. **The module handbook.** Course-level rules override general ones and are frequently stricter or more specific.
2. **The assessment brief for this piece of work.** The most specific rule wins, and briefs sometimes carry a line that contradicts the general policy.
3. **The academic integrity or academic misconduct policy**, usually on the university's registry or quality-assurance pages.
4. **Any separate generative-AI guidance**, which many institutions issued from about 2023 onwards and revise often.

Download them. Policies get updated, and the version that applied when you submitted is the one you want to be able to produce.

The three-tier structure most policies use

Despite enormous variation in wording, most institutions have converged on some version of three categories, often called a traffic-light or scale approach:

- **Not permitted.** No AI at any stage. Common for language assessments, some exams, and any task where the thing being tested is exactly what a model does.
- **Permitted with disclosure.** Use it, say how. The largest category and growing.
- **Permitted, encouraged, or required.** Increasingly common in professional and technical subjects where using the tools is part of the competence being assessed.

A well-written brief tells you which tier applies. If it does not, that is the question to ask.

Words that matter, and what they usually mean

Policies are written carefully and read carelessly. The distinctions that decide cases:

- **"Generate" versus "assist".** Producing new text for your submission versus improving text you wrote. Almost every policy treats these differently, and it is the main line.
- **"Substantial" or "significant".** Usually undefined. Ask what it means for your assessment; the answer in writing is worth having.
- **"Proofreading" and "language support".** Usually permitted, sometimes with a disclosure requirement, and sometimes limited to correcting errors rather than rewriting.
- **"Your own work".** The oldest phrase in the document, and the one everything else is interpreting.
- **"Undisclosed".** Several policies make non-disclosure the offence rather than the use. Under those, a permitted use becomes misconduct simply by not being declared, which is a trap for the honest.

When the policy is silent or unclear

This is common, and the answer is the same every time: **ask in writing, and keep the reply.**

> *Dear Dr X, for the coursework due on the 14th, could you confirm whether it is acceptable to use an AI assistant to (a) generate practice questions while revising, (b) check my draft against the mark scheme, and (c) correct grammar in my own prose? I want to make sure my disclosure statement is accurate.*

Three things this achieves. You get an answer. You have it in writing, which is evidence. And the framing signals that you intend to disclose, which changes how the question is read.

Email, not a corridor conversation. A remembered conversation is not evidence and honest people misremember.

The safe default when you cannot find out

If you genuinely cannot establish the rule before a deadline:

- Use it only for things that produce no text in your submission — practice questions, explanation, structural diagnosis of your own draft.
- Write every word yourself.
- Disclose what you did, briefly and factually.

That combination is inside almost every policy in existence. It is not the most efficient possible use; it is the one that cannot go wrong.

Policies vary more than you would expect

Between countries, between institutions, and between departments in one institution. A computing department may require AI use; the law school next door may prohibit it entirely; the language centre may prohibit it for one assessment and encourage it for another. All of that can be true simultaneously and consistently, because they are assessing different things.

So do not import a rule from another course, another university, or the internet. The document that applies to you is the one for your module, and the fact that a friend on another programme does something different is not evidence about your position.

And the honest part

These policies are new, they are being revised constantly, and many of them are internally inconsistent or unenforceable as written. Institutions are working it out in public. That is a real situation and not a reason to ignore the rules — it is a reason to document what you did, ask when unsure, and disclose generously. Those three behaviours protect you under any version of a policy, including the one your institution has not written yet.

BEFORE YOU MOVE ON

Why is a policy that makes non-disclosure the offence a particular hazard for honest students?

A. Because it shifts the burden of proof onto the student to show they did not use AI, which is impossible.

B. Because it requires students to declare uses that other policies would treat as ordinary tool use, which is easy to overlook.

C. Because it means any use of AI is prohibited unless permission was sought in advance.

D. Because disclosure statements are read by markers who may then judge the work more harshly, so honest students are penalised for candour.

58 · 8 min

Writing a disclosure that helps you

THE ONE THING TO KEEP

A disclosure is a factual account, not a confession — name the tool, the stage, what happened to the output, and state that you obtained and read every source, because unread sources is the actual offence in most cases.

Specific beats vague, always

A disclosure statement is a factual account of what you did. Vague statements make markers suspicious; specific ones close the question.

Compare:

> *AI was used in the preparation of this assignment.*

against

> *I used a language model in two ways. Before drafting, I asked it to generate practice questions on the reading to test my understanding; none of that output appears here. After drafting, I pasted my complete draft and the mark scheme and asked which criteria were weakest; I acted on two of its four suggestions by rewriting my third and fifth paragraphs myself. All prose, all argument and all source selection are my own, and I obtained and read every source cited.*

The first invites a question. The second answers it. It takes ninety seconds to write and it is the cheapest protection available.

What to include

- **Which tool**, and roughly when.
- **What you asked it to do**, by stage — before, during, or after drafting.

- **What happened to the output.** Used, discarded, acted on by rewriting yourself.
- **What is unambiguously yours.** Argument, structure, prose, source selection.
- **The verification claim.** That you obtained and read every source.

That last line matters more than the rest combined, because unread sources is the actual offence in most cases.

Three worked examples

Minimal use.

> *AI use: I used a language model to check spelling and grammar in prose I wrote, and to generate revision questions while preparing. No generated text appears in this submission. All sources cited were obtained and read.*

Substantial permitted use.

> *AI use: I used a model to explain two methods sections I could not follow (papers 4 and 7 in the reference list), to suggest which statistical test suited my design — the choice and the justification in section 3.2 are mine — and to critique the structure of my draft. The analysis was run in jamovi. All writing is my own; all sources were obtained and read.*

Code.

> *AI use: I used an assistant to explain error messages and to explain an unfamiliar library function. I wrote all submitted code and all tests. Two functions were revised after the assistant identified a boundary condition I had missed; the revisions are mine. Commit history is in the repository.*

Notice all three are short, factual, and unapologetic. A disclosure is not a confession.

Where to put it

Where the brief says. If it does not say: a short paragraph after the references, headed *Statement on AI use*. In a repository, in the README and referenced in the submission.

If your institution provides a form or a declaration box, use theirs and do not substitute your own wording.

The thing a disclosure cannot do

It does not make a prohibited use permitted. Declaring that you generated the essay does not turn generating the essay into acceptable practice under a policy that forbids it; it makes the finding faster.

It also does not cover what it does not mention. A disclosure that omits a use you made is worse than none, because it is now an inaccurate statement rather than an absent one, and that changes the character of the matter substantially.

So: disclose everything you actually did, including the things you feel slightly uncomfortable about. If a use is uncomfortable to write down, that discomfort is the information — it is telling you something about the use, and the moment to act on it is before submission.

Disclose generously

When in doubt, include it. There is no penalty anywhere for declaring a permitted use, and there is a real penalty for omitting one that mattered. Over-disclosure costs you two extra sentences.

And write it as you go, not at the end. A note in your working file each time you use a tool takes five seconds and means the statement is accurate rather than reconstructed from memory the night before, when everybody forgets the thing they did in week two.

The test of a good disclosure

Could your marker, reading it, tell exactly what you did without asking you a follow-up question? If yes, it is sufficient. If they would need to ask *but did you write the prose*, it is not, and you should add the sentence that answers it.

BEFORE YOU MOVE ON

Why is a disclosure that omits one of the uses you made worse than writing no disclosure at all?

A. Because incomplete disclosures are treated as attempts to conceal, which carries a heavier penalty than the underlying use.

B. Because markers compare disclosures against detector output, and a mismatch triggers an investigation.

C. Because it becomes an inaccurate statement rather than an absent one, which changes the character of the matter.

D. Because a partial disclosure implies the omitted use was the one the student knew to be prohibited, which is how a marker will read it.

---

59 · 9 min

How detectors work, and why they cannot

THE ONE THING TO KEEP

Detectors measure perplexity and burstiness — how typical the writing is — so they systematically misclassify second-language and formulaic writers, and the base rate means roughly half of flagged essays can be innocent even with generous accuracy assumptions.

What they actually measure

AI text detectors do not identify the origin of a document. They measure statistical properties of the text and compare them against what generated text

tends to look like. The two most common signals:

Perplexity. Roughly, how surprised a language model is by each word given the previous ones. Generated text tends to have low perplexity, because it was produced by picking likely continuations. So a detector infers: unsurprising text, probably generated.

Burstiness. Variation in sentence length and complexity across a document. Human writing tends to vary more; generated writing tends to be more even.

Now read those again with a student in mind. What kind of human writing has low perplexity and low burstiness? Careful, formal, well-edited prose in a conventional structure. A student writing to a rubric in a second language, using standard academic phrasing, produces exactly that profile.

The detector is measuring how typical your writing is, not who produced it. That is not a flaw in a particular product. It is what the measurement is.

The base rate problem

Suppose a detector is 95 per cent accurate in both directions — better than the evidence supports — and 5 per cent of a 1,000-essay cohort used AI in a prohibited way.

- 50 essays are AI-written. 47 or 48 are flagged.
- 950 essays are human. 5 per cent of them, about 47, are also flagged.

So about half of all flagged essays are innocent. That is with generous accuracy assumptions and a fairly high prevalence. Lower the prevalence or the accuracy and it gets worse fast.

A detector that is 95 per cent accurate, on a thousand essays

	Flagged	Not flagged
Used AI (50)	**47** correctly caught	**3** missed entirely
Did not (950)	**47** wrongly accused	**903** cleared

Take the generous case and assume one essay in twenty broke the rule. Forty-seven of the fifty are caught, and so are forty-seven of the nine hundred and fifty who did nothing. About half of everyone flagged is innocent, and lowering either the accuracy or the prevalence makes it worse quickly. This is the arithmetic that governs medical screening, and a screening result is not a verdict.

This is the same arithmetic that governs medical screening, and it is why a positive result on a rare condition often means very little on its own. A detector score is a screening result, and treating it as a verdict is a category error with a name.

The unequal harm

The false positives are not randomly distributed, and this is the part that should trouble everyone. Published evaluations have found detectors flagging writing by non-native English speakers at substantially higher rates than native speakers — in one widely-reported study, a majority of essays by non-native speakers were misclassified as AI-generated, while nearly all native-speaker essays were classified correctly.

The mechanism explains it exactly. Learners of a language use more common words, more standard constructions, less varied sentence structure. That is low perplexity and low burstiness. The detector is measuring the language proficiency and reporting it as authorship.

The same logic exposes autistic students who write in a consistent register, students taught to write formulaically, and anyone using an editing tool — including the spelling and grammar checkers that have been permitted for decades.

What the vendors and institutions now say

Several major providers state that their scores are indicative rather than conclusive and should not be the sole basis of an allegation. OpenAI withdrew its own classifier in 2023, citing low accuracy. A number of universities have disabled detector tools in their submission systems for exactly these reasons, and others have adopted policies stating that a detector score alone cannot support a finding.

So if a score is being used against you as though it were proof, that is not only unfair — it is usually contrary to the tool's own documentation, and saying so calmly is a legitimate part of a response.

What this does not mean

It does not mean nobody can tell. Markers who know a subject and have read your previous work recognise a change in register, an argument that does not connect to the seminar you were in, a bibliography with fabricated entries, and an answer that cannot be defended in conversation. Those are evidence, and they are how most cases are actually made and how most are correctly made.

It also does not mean generated submissions go undetected in practice. The most reliable methods are the human ones: a viva, a question about your own paragraph, an in-person written task compared against your coursework.

What to do about it

You cannot control whether a detector flags you. You can control whether you have evidence, and that is the whole of the next lesson.

- Write in a tool with version history and leave it on.
- Keep drafts, notes, and the reading you did.
- Keep the unaided-check log from Module 3.
- If your writing is naturally formulaic, or you are writing in a second language, be especially deliberate about this. The system's error falls on you disproportionately, and the only available protection is a record.

That is an unjust distribution of effort. It is also, currently, the situation.

BEFORE YOU MOVE ON

A detector is 95 per cent accurate in both directions and 5 per cent of a 1,000-essay cohort used AI. Why does a flag tell a marker so little?

A. Because 95 per cent accuracy is measured on benchmark datasets that do not resemble student coursework.

B. Because a 5 per cent prevalence estimate is itself a guess, and without knowing the true rate no inference can be drawn from a positive result.

C. Because accuracy figures describe the tool's performance on average and cannot be applied to an individual essay at all.

D. Because the innocent group is so much larger that its false positives roughly equal the true positives.

---

60 · 8 min

The process record

THE ONE THING TO KEEP

A process record cannot be created after an accusation, which is exactly why it works — write in a tool with version history on, never paste a finished draft into an empty document, and photograph paper drafts.

Evidence you cannot make later

The strongest defence against any allegation about your work is a record of it being made, and it has one property that decides everything: **it cannot be created afterwards.** Version history is timestamped by the platform. Commits carry dates. Notes accumulate. None of it can be produced the week after an accusation, which is exactly why it is convincing.

So the record has to exist before you need it, which means it has to be a habit rather than a project.

The cheapest possible version

One rule: **write in a tool that keeps version history, and never paste a finished draft into an empty document.**

Google Docs keeps detailed version history automatically and free. Microsoft Word with autosave to OneDrive or SharePoint does the same. Both let you show a document growing over hours and days, with pauses, deletions, and reordering — the shape of somebody writing.

That single habit is most of the protection, and it costs nothing.

The common way people destroy their own evidence is drafting somewhere else and pasting the result in. A document whose history shows 2,000 words arriving in one action at 23:40 tells a story you cannot easily contradict, even when you wrote every word in another window.

What a good record contains

For a substantial piece of work:

- **Version history**, on, from the first sentence.
- **Your reading notes**, in your own words, dated, with page numbers.
- **The outline**, as it changed. The old versions are the valuable ones because they show thinking.
- **Your search history for sources** — a Zotero library with items added over weeks is a beautiful record.
- **Any conversations** with an assistant that you would disclose. Keep the links or exports.
- **The disclosure statement**, written as you went.
- **Your unaided-check log** from Module 3, which shows capability developing independently of any tool.

For code: git, with commits at the end of every session, including the failing ones. A history with forty commits showing an idea developing, with dead ends in it, is far more persuasive than a repository containing one commit called *final*.

The failure that looks bad and is innocent

Worth naming so you can avoid it. Students who draft on paper, or in a notes app, or in an assistant's window, and then type or paste the result into the submission document, produce a history with no development in it. They did the work honestly and the record shows nothing.

If you genuinely prefer drafting on paper — many people write better that way — photograph the pages, dated. Two minutes, and your record is intact.

Do not fake it

There are tools that simulate typing to fabricate a document history. Using one is a separate and more serious offence than whatever it was meant to conceal: it is deliberate fabrication of evidence, and institutions treat it accordingly. It is also detectable, because simulated typing has a suspiciously even rhythm and no deletions, revisions or long pauses.

A real writing history is messy. That messiness is the signal.

The record helps you before it ever defends you

The framing that makes this bearable: none of this is extra work done out of fear. Every item on that list is something that makes the work better.

Notes in your own words help you write. An outline that changed is an argument that developed. Commits at the end of each session let you go back when you break something. A disclosure written as you go is accurate. The unaided-check log tells you what you can actually do.

You are keeping a record because it is how good work is done. That it also happens to be unanswerable evidence is a side effect, and it is the reason to start in week one rather than the week you need it.

Retention

Keep everything until your degree is awarded and the appeal window has closed. Cases sometimes arrive months later, and a cloud document deleted in a tidy-up is gone. Export a copy of anything that matters at the end of each year — institutional accounts are closed after you graduate, sometimes with little notice.

BEFORE YOU MOVE ON

A student writes an entire essay by hand, then types it into a shared document in one sitting. Why is their record weak despite the work being entirely their own?

A. Because the version history shows a finished text arriving rather than a document developing.

B. Because handwritten drafts are not accepted as evidence in academic misconduct proceedings.

C. Because typing up a completed draft produces the even rhythm that fabricated histories also show.

D. Because the timestamps will cluster near the deadline, and late-clustered activity is treated as a warning sign in itself.

61 · 8 min

If you are accused

THE ONE THING TO KEEP

An allegation is not a finding — do not confess to what you did not do, get your students' union adviser involved before responding at length, and remember that being able to discuss your own work in detail is the strongest evidence there is.

First, do not panic and do not confess

An allegation is not a finding. Procedures exist, they involve a hearing or an interview, and most institutions have an appeal route afterwards. People are cleared regularly.

The single most damaging thing students do is admit to something they did not do, because the process is frightening and admitting seems like the fastest way out. It is not. A finding of misconduct on your record is a serious and lasting thing, and it is much harder to undo than to contest in the first place.

Equally: do not lie. If you did something, saying so early is treated far more leniently everywhere, and being caught in a false account is worse than the original matter.

The steps, in order

1. Read the allegation carefully. What exactly is alleged? Generated text? A fabricated citation? Undisclosed use? These are different charges with different evidence and different responses. Students often defend against the wrong one.

2. Find out what the evidence is. You are normally entitled to know. If it is a detector score, ask for the score, the tool, and the institution's policy on how such scores may be used.

3. Gather your record. Version history, notes, outline drafts, reading, commits, browser history, library loans. Export it now, before anything is archived or an account is closed.

4. Get support before you respond. Your students' union almost certainly has advisers who do this every week, free, and they know your institution's procedure and its people. This is what they are for and they are chronically underused. Many institutions also have an independent student advice service.

5. Do not respond alone at length in writing before you have taken advice. A short acknowledgement that you have received the allegation and are preparing a response is fine.

6. Prepare to discuss the work. The strongest evidence you have is being able to talk about your own essay: why you argued what you did, why the third paragraph comes third, what the sources say, what you rejected and why. Somebody who did the work can do this and somebody who did not cannot. Re-read your own work before any meeting.

Responding to a detector score

Calmly, factually, and without a lecture. The points that matter:

- Detector output is probabilistic and vendors state it is indicative rather than conclusive.
- Documented false positive rates are materially higher for second-language writers and for formulaic prose. If that applies to you, say so.
- Ask what corroborating evidence exists beyond the score, and what the institution's policy says about scores as sole evidence.
- Offer your process record, and offer to discuss the work.

The tone matters. You are not arguing that detection is impossible; you are pointing out that this particular instrument cannot support this particular conclusion on its own, and offering better evidence.

If you did use it in a way you should not have

Say so, early, plainly, without elaborate explanation. Every institution's procedure treats early admission more leniently, and the alternative — a contested case that goes against you — carries a worse outcome and a worse record.

And if you realise it *before* submission or before anyone notices, tell the module leader yourself. Self-disclosure before detection is treated differently from being caught, at every institution.

The unfair situation, named

Some students will be accused wrongly. The detector arithmetic guarantees it. If it happens to you it will be one of the more distressing experiences of your degree, and knowing that the instrument is unreliable does not make the meeting less frightening.

Three things reduce the damage, and all three have to be in place beforehand: a process record, a habit of disclosure, and the ability to discuss your own work in detail. That is why this module comes before the exam module rather than after it, and why the record habit is worth building in your first week rather than the week you need it.

Afterwards

If you are cleared, ask for confirmation in writing and check that nothing remains on your record. If a finding is made and you believe it wrong, use the appeal route; there is a deadline and it is usually short. Either way, keep every document from the process.

BEFORE YOU MOVE ON

Why is the ability to discuss your own essay in detail stronger evidence than any document you can produce?

A. Because documents can be fabricated whereas a conversation cannot be prepared for in advance.
B. Because panels weight oral evidence above documentary evidence in academic misconduct procedures.
C. Because someone who made the argument can explain why each choice was made, and someone who did not cannot.
D. Because it demonstrates subject knowledge, and a student who knows the subject would have had no reason to use AI in the first place.

62 · 8 min

Where the line really is

THE ONE THING TO KEEP

Ask what the assessment is trying to measure and whether your use interferes with measuring it — a test that generalises to tools that do not exist yet, unlike any list of rules.

A spectrum, not a switch

People argue about AI and cheating as though there were one line. There is a spectrum, and most of it has been settled for decades under other names.

Roughly from unproblematic to prohibited:

1. **Spellcheck.** Universal, permitted everywhere, nobody discusses it.
2. **Grammar checking.** Long-standing, permitted almost everywhere.
3. **A thesaurus or dictionary.** Ancient, permitted.
4. **Translation of a source you are reading.** Normal scholarly practice.

5. **Explanation of a concept.** Identical to asking a tutor, a friend, or reading a textbook.
6. **Generating practice questions.** Produces nothing in the submission. Safe everywhere.
7. **Critique of your draft.** Identical to a writing centre appointment, which institutions provide and encourage.
8. **Editing your prose at sentence level.** Where policies begin to differ, and where the *correct an error* versus *rewrite it for me* distinction starts to matter.
9. **Translation of your own writing into the assessed language.** Usually prohibited in language assessments, variable elsewhere, and frequently requires disclosure.
10. **Generating text you submit.** Prohibited in most current policies unless explicitly permitted.
11. **Generating the whole submission.** Prohibited everywhere it is not the point of the exercise.
12. **Fabricating sources or data.** Serious misconduct, and it always was.

Items 1 to 7 are, for almost every policy in the world, fine. That is a great deal of useful help, and students who believe the whole category is forbidden are denying themselves things their institution actively provides.

Settled, and not settled

Fine under almost every policy

- Spellcheck, and grammar checking
- A dictionary, a thesaurus, a translation of a source you are reading
- Explanation of a concept, which is what a tutor or a textbook does
- Generating practice questions, which produces nothing you submit
- Critique of your own draft, which is what a writing centre does

Where policies differ, or prohibit

- Editing your prose at sentence level, where correcting meets rewriting
- Translating your own writing into the assessed language
- Generating text that you submit
- Generating the whole submission
- Fabricating sources or data, which was always serious misconduct

Students who believe the whole category is forbidden deny themselves things their institution actively provides and pays for. One question resolves most of the rest: what is this assessment trying to measure, and does my use interfere with measuring it? That generalises to tools which do not exist yet, and no list of rules does.

The question that resolves most cases

> ***What is this assessment trying to measure, and does my use interfere with measuring it?***

A translation exercise measures your ability to translate; using a translator defeats it entirely. An essay measures your ability to construct an argument; having grammar corrected does not touch that, and having the argument constructed destroys it. A programming assignment measures whether you can write the program; having an error message explained is fine, having the function written is not.

This single question resolves most cases correctly, and it generalises to tools that do not exist yet, which no list of rules does.

The two-part test

When the question above is still unclear, apply both parts:

Would I be comfortable writing exactly what I did in the disclosure statement? If not, that discomfort is data. It is usually the most reliable signal you have, and it arrives before the deadline.

Could I now do this unaided? If a use leaves you able to do it yourself next time, it taught you something. If it leaves you unable, it substituted for you. This is the test that protects your education rather than your record, and it is the one that matters in five years.

Three cases worth thinking through

A student asks a model to explain a concept, understands it, then writes the essay entirely themselves. Fine everywhere. Identical to reading a textbook. No disclosure required under most policies, and disclosing costs nothing.

A student writes an essay, asks a model to improve the prose, and accepts most of the changes. Depends on the policy and on how much changed. If the argument, structure and evidence are yours and the sentences

have been smoothed, most policies permit it with disclosure. If entire sentences now say things you did not write, you have crossed into 10. The honest test is whether you can explain every change.

A student uses a model to translate their essay from their first language into English for submission. Genuinely contested. Some institutions treat it as language support and permit it with disclosure; others treat it as generating the submitted text. It also raises a fairness question, since a student writing in their first language gets a benefit their classmates do not need. Find out your institution's position before submitting rather than after, and if there is not one, ask.

The part nobody has settled

Where the line should sit is a live argument, and it is being argued differently in different countries, disciplines and institutions. Some academics think most of this will be permitted within a few years and assessment will change to measure other things. Others think closed assessment will expand. Both are defensible; neither is settled.

What is settled, and has been for a very long time: submitting work as your own when it is not, and citing sources you have not read. Those two hold in every version of the future being argued about, and they are the two that generate almost all of the actual cases.

BEFORE YOU MOVE ON

Why does "what is this assessment measuring, and does my use interfere with measuring it?" work better than a list of permitted tools?

A. Because it is the criterion institutions actually apply when deciding cases, so it predicts outcomes accurately.

B. Because it generalises to tools that do not exist yet, which no list can do.

C. Because it removes the need to read the policy, since the underlying purpose of assessment is the same everywhere.

D. Because it produces a stricter standard than most written policies, so a student applying it will always be within the rules.

63 · 8 min

Accessibility, fairness and who these rules fall on

THE ONE THING TO KEEP

An adjustment removes a barrier to demonstrating what you know without supplying what you know — get it documented formally, because a registered adjustment is a complete answer rather than an explanation offered afterwards.

Legitimate adjustment is not cheating

For some students these tools are not a shortcut; they are the thing that makes the work possible on equal terms.

- A dyslexic student using text-to-speech and grammar support is doing what assistive technology has always done, and most institutions have provided exactly these tools for years.
- A student with a motor impairment dictating instead of typing is changing the input method, not the authorship.
- A student with ADHD using a model to break a task into steps is using a structuring aid.
- A blind student having a figure described is accessing material the sighted student can already see.
- A second-language student checking whether a sentence is idiomatic is levelling something that was never part of what the essay was meant to measure.

None of these substitutes for the thinking. All of them remove a barrier that is not the subject of the assessment.

Get it on the record

If you have a disability or a specific learning difference, register with your institution's disability service and get the adjustments documented. This matters for three reasons.

It makes the adjustment official rather than something you are quietly doing. It usually unlocks tools and support, often including paid software you would otherwise not have. And if any question is ever raised about your work, a documented adjustment is a complete answer rather than an explanation you are offering after the fact.

The process is often slow and involves paperwork. Start it early in the year, not in the week of a deadline.

The uneven distribution of risk

The rest of this lesson is uncomfortable and it is better said than not.

The costs of the current arrangements do not fall evenly. Detectors misclassify second-language writers at markedly higher rates, so students already working in a language not their own carry an additional risk their classmates do not. Students who were taught to write in a formulaic structure — which is what a great many school systems teach, particularly where exams reward a fixed essay shape — produce exactly the low-variation prose that detectors flag. Students without money use free tiers and free tools; students with money use better models and, in some cases, entirely undetectable arrangements. Students without a strong command of the institution's procedures are less likely to challenge an allegation and more likely to accept a finding.

So the current system is more likely to catch the least advantaged and less likely to catch the most. That is not an argument for using AI dishonestly. It is a reason to be sceptical of confident claims about fairness in this area, and a reason for the protective habits in this module to be treated as more urgent by exactly the students who have least reason to expect the benefit of the doubt.

What to do about your own position

If any of the categories above describes you:

- Keep the process record, without exception. Version history on, always.
- Disclose generously, including things you think are obviously fine.
- Register any adjustment formally.
- Know where your students' union advice service is before you need it.
- Keep samples of your writing from across the year, so a change in register has an innocent explanation with evidence attached.

None of that is fair as a distribution of effort. It is what reduces the risk under the arrangements that exist.

And the argument worth making

Institutions are actively revising these policies, and student input has changed several of them. If your institution uses detector scores as sole evidence, that is contrary to what the vendors themselves say, and it is a reasonable thing for a students' union to raise. If your course prohibits tools that its own disability service provides, that inconsistency is worth naming.

The rules are being written now, in a hurry, by people who are also working it out. Being the student who reads the policy and asks a precise question about it is a more useful contribution than it sounds.

The thing that does not change

An adjustment removes a barrier to demonstrating what you know. It does not supply what you know. A dyslexic student using text-to-speech still has to have read and understood the material; a dictating student still has to have the argument. If a use is supplying the substance rather than the access, it is not an adjustment, whatever it is called — and that distinction is one you can apply honestly to your own case better than anybody else can.

BEFORE YOU MOVE ON

What distinguishes a legitimate accessibility adjustment from a use that crosses into misconduct?

A. Whether the student has a formally registered disability, since registration is what makes the same tool permitted.
B. Whether the tool was provided by the institution, since institutionally supplied software is by definition permitted.
C. Whether the assessment criteria include the skill the tool replaces, since an adjustment is only legitimate where the barrier is incidental to what is being marked.
D. Whether it removes a barrier to demonstrating knowledge, or supplies the knowledge itself.

64 · 8 min

What happens to what you type

THE ONE THING TO KEEP

Storage, training and human review are three separate questions with different answers — the training toggle is usually yours to set, an institutional account is a work account with logs, and the only content safe from every future is content you did not type.

Three different questions

Students collapse three separate things into one worry. Separating them makes each answerable.

Is it stored? Almost always yes, for some period. Providers retain conversations for abuse monitoring and service operation even when you delete them from your view. Retention periods are stated in the privacy policy and typically run to some number of days after deletion.

Is it used for training? This is the one you can usually control. Consumer free tiers commonly train on input by default with an opt-out; paid business, enterprise and education tiers commonly do not, by contract. The setting is in data controls and is worth finding once.

Can a human read it? Sometimes, in limited circumstances: samples reviewed for safety, investigations of abuse, legal process. Not routinely, and not casually.

Your institution's account

If your university provides a tool, the contract usually prohibits training on your input and places the data under an agreement the institution has assessed. That is a real advantage and it is often a paid tier at no cost to you.

The trade-off: logs exist and are accessible to the institution under its own procedures. Administrators are not reading your conversations for entertainment, and in an investigation the records may be available. Treat an institutional account as a work account, because that is what it is.

What a school or university can see anyway

Broader than most students assume, and worth knowing accurately rather than fearing vaguely:

- **On their network**, traffic to and from your device is visible in outline, and on a managed device more than that.
- **On a managed laptop or Chromebook** issued by the institution, potentially a great deal, including installed monitoring.
- **In their systems** — the virtual learning environment, submission platform, email — everything, and it is retained.
- **Submission platforms** store your work and compare it against their databases, which is what they exist for.

On your own device on your own connection, they see essentially nothing.

Age, and school students

If you are under 18, additional protections apply and additional restrictions come with them. Most providers set minimum ages, often 13 with parental consent up to 18, and school-provided tools usually run under stricter configurations with filtering and logging. Data protection law in many countries — GDPR in Europe and the UK, COPPA in the United States, and equivalents elsewhere — treats children's data more restrictively.

The practical version for a school student: use the tool your school provides if there is one, assume anything typed into a school account is visible to the school, and do not put personal details about yourself or your friends into any of it.

Your own data, and the settings pass

Once per tool, ten minutes:

1. **Data controls.** Find the training toggle. Decide.
2. **History and retention.** Learn what deleting does and how long anything persists.
3. **Memory.** Read what it has stored about you. People are regularly surprised, and it is editable.
4. **Connected apps and integrations.** Anything with access to your email, files or calendar — check what you granted and remove what you do not use.
5. **Export.** Most providers let you download everything they hold. Doing it once is instructive.

Rights you have

In many jurisdictions, data protection law gives you rights over your personal data: access to what is held, correction, deletion in some circumstances, and portability. The GDPR versions apply across the EU and UK; several other countries have similar regimes, India's data protection legislation among them, and the details differ.

The practical point is that these are real and usable — a subject access request is a formal thing you can make of a provider or your institution — and that the specifics vary by country enough that this course is not going to summarise them into advice. Look up your own jurisdiction's regime if you need to use it, and none of this is legal advice.

The habit that makes all of this simpler

The rule from Module 2, restated because it is the whole of the practical answer:

> *Would I mind if this appeared in a public archive with my name on it?*

Settings are promises about handling. Companies change, are acquired, are breached, and are subject to legal process. The only content genuinely safe from all of that is content you did not type. So paste generously from published and your own material, keep other people's information out entirely, and do the settings pass once so that the remaining risk is one you chose rather than one you never looked at.

BEFORE YOU MOVE ON

A student turns off the training setting on their account. What does this change, and what does it not?

A. It stops storage entirely, so conversations are no longer retained after deletion.

B. It stops human review, since review exists to support training data quality.

C. It stops their input improving the model, but the content is still stored and can still be reviewed in limited circumstances.

D. It applies retroactively as well as prospectively, so material already used in training is removed from the model at the next update.

CASE STUDY · MODULE 7

Nineteen flags and four days

A module leader at a university in Hyderabad with 240 first-year essays, a detector report, and a memo from the head of department asking for firm action.

The submission platform's AI indicator had been switched on for the first time that session. Of Dr Imran Qureshi's 240 first-year essays, nineteen came back above the department's threshold of forty per cent.

He read the list before he read any of the essays, and the first thing he noticed was who was on it. Fifteen of the nineteen were students whose first language was not English. About fifty-five per cent of the cohort were second-language writers, so the flags were falling on that group at roughly twice the rate they fell on everybody else.

He had three weeks until marks were due, five other modules' worth of marking behind them, and a memo from the head of department asking for firm and visible action after a case elsewhere in the faculty had gone badly and publicly.

Before doing anything he worked the arithmetic on a sheet of paper, because it is the kind of arithmetic that is easy to assume and hard to guess. Assume the tool is ninety-five per cent accurate in both directions, which is more generous than the published evidence supports. Assume one essay in twenty in his cohort broke the rule. That gives twelve rule-breakers among 240, of whom eleven or twelve are flagged, and 228 innocent essays of which about eleven are also flagged. So roughly half of anybody flagged has done nothing at all. Drop the accuracy or the prevalence and it gets worse quickly, and both are more likely to be lower than higher.

That ruled out the fastest option. Referring all nineteen on the score would have looked decisive, would have satisfied the memo, and would have been contrary to what the vendor's own documentation says about its output being indicative rather than conclusive. It would also have put about nine innocent first-years through a misconduct process in their first semester, dispropor-

tionately students already carrying the work of writing in a language not their own.

Referring none was not available either, and he was clear with himself about why. If one of the nineteen had generated an essay and he had looked away on the grounds that the instrument is unreliable, that is his failure, and the unreliability of the instrument does not transfer the responsibility anywhere.

The third option cost him exactly the thing he did not have. Fifteen-minute conversations with each of the nineteen, with preparation, is about eight hours across four days. It delays the marks of all 240 students, most of whom had done nothing and were waiting. And being summoned to discuss your own essay is frightening however carefully the email is worded, so the process itself imposes a real cost on people who will turn out to have done nothing wrong.

He took the third option and changed one thing about it, which is the decision the case turns on. He added eleven essays chosen at random from the unflagged 221 and invited all thirty students to a conversation about their work, described accurately as a sampling exercise. Being asked was no longer an accusation, because being asked no longer meant anything about the score. He also put the references from all thirty essays through the DOI resolver and the library catalogue, which took one afternoon.

STOP HERE

Before turning the page: what would you have done, and what would it have cost if you were wrong? Write it down. The point of the next section is to compare.

WHAT HAPPENED

Three of the nineteen could not discuss their own essays at all: not the argument, not why the third section came third, not what any cited source said. Two of those three had references that did not resolve.

One of the eleven randomly chosen unflagged essays had two references that did not exist, and its author could not say what either paper argued.

Sixteen of the nineteen discussed their work without difficulty, and eleven of those sixteen were second-language writers. Two showed him Google Docs histories running to more than twenty sessions.

One of them, Hamza, had a version history that looked far worse than anybody he referred: two thousand four hundred words appearing in a single action at ten past eleven at night. He drafts on paper, which many people do, and types the result in. What saved him was that a teacher in his first week had told him to photograph the pages, and he had eleven dated photographs of a notebook with crossings-out, an abandoned second argument and a paragraph written twice.

Dr Qureshi referred four students in total, on the evidence of fabricated references and the conversations. The detector score is named nowhere in any of the four case files. Three findings were made and one student was cleared.

A term later the department turned the indicator off in the submission platform and replaced it with a random ten per cent reference audit, which takes a fraction of the time and finds something a process can actually stand on.

One thing did not resolve. Several of the fifteen second-language students who were flagged and cleared changed how they wrote afterwards. One told him she had started deliberately varying her sentence lengths. His view is that a system which teaches a careful student to write worse in order to look more human has failed at something more important than detection.

Work his arithmetic and say what the nineteen flags could support on their own.

With ninety-five per cent accuracy in both directions and five per cent prevalence in 240 essays, about twelve students broke the rule and eleven or twelve of them are flagged, while about eleven of the 228 who did nothing are also flagged. Around half of everyone flagged is innocent, and those assumptions are more generous than the published evidence supports. So the flags can support one thing only: a decision about where to look next. They are a screening result, and a screening result for an uncommon condition carries little information on its own, which is the same arithmetic that governs medical screening. Treating it as a verdict is a category error, and in this case it is also contrary to what the vendors themselves say their scores mean.

Hamza's version history looked worse than that of the students who were referred. Explain why, and say what made it survivable.

Version history records the shape of a document arriving, not the shape of work being done. Somebody who drafts in another place — on paper, in a notes app, in a chat window — and types or pastes the result produces a history with no development in it, which is indistinguishable from the history of somebody who generated the text elsewhere. He had done nothing wrong and his strongest evidence looked like the worst evidence in the room. What saved him was a second record that could not have been made afterwards: dated photographs of notebook pages carrying crossings-out, an abandoned argument and a paragraph written twice. Messiness is the signal, because it is the thing nobody produces when reconstructing a history after an accusation.

A student who was cleared now varies her sentence lengths deliberately. What is the instrument measuring, and what is she optimising for?

It measures statistical typicality — roughly, how predictable each word is given the ones before it, and how much sentence length and complexity vary across the document. Low perplexity and low burstiness are what generated text tends to show, and they are also what careful, plain, formally structured writing by a learner of the language shows, which is why the false positives fall where they do. She is now optimising for burstiness rather than for clarity. Everything that makes prose easy to read pushes the score the wrong way, so the instrument is asking her to trade a real quality for an appearance. That is not a flaw in one product to be patched in the next version; it is what the measurement is, and it is why the defensible protections are a process record and a conversation about the work.

MODULE 7 · TEST

Check what stuck

6 questions, one right answer each. This one is for you, not for a record: answers and the reason behind each are at the back. If two questions from the same module go wrong, re-read that module before the exam rather than after.

1 What does an AI text detector actually measure?

A. A watermark embedded by the model that produced the text
B. Whether the phrasing appears in the vendor's output database
C. How typical the writing is, not who produced it
D. The proportion of sentences matching a known training corpus

2 A detector is 95 per cent accurate in both directions, and one essay in twenty out of a thousand broke the rule. Roughly what share of flagged essays are innocent?

A. About half
B. About five per cent, matching the stated error rate
C. Almost none, because the accuracy is so high
D. About a quarter, since most of the flags are correct

3 Which line in a disclosure statement does the most work?

A. A sentence naming the model version and the date of use
B. That you obtained and read every source you cite
C. A word count of any text the model produced for you
D. An apology for having used assistance on assessed work

4 Which habit produces a version history that looks bad despite the work being entirely honest?

A. Committing code at the end of every session, including bad ones
B. Keeping an outline that changed several times during the work
C. Writing the disclosure statement as you go rather than at the end
D. Drafting on paper and typing the finished result in

5 Which question resolves most cases about whether a particular use is acceptable?

A. Would my marker be able to tell that I had used a tool?
B. Is this use listed on the university's public policy page?
C. Does my use interfere with what this assessment measures?
D. Have other students on the same module done the same thing?

6 Which students are most likely to be wrongly flagged, and why?

A. Students who write in long, varied sentences with rare words
B. Students taught to write plainly and in a second language
C. Students who submit late and are marked by a different tutor
D. Students who use a reference manager to format their citations

MODULE 8

Assessment, and the years after it

Past papers, timed practice, vivas, long projects, applications, and the honest question of what is worth learning when the tools can do a great deal of it.

> BY THE END OF THIS MODULE YOU CAN
>
> Prepare for closed and oral assessment by practising under the conditions of the test, manage a long project with a defensible record, and decide which capabilities to build unaided given what the tools can and cannot do

65 · 8 min

Revision That Survives The Exam Hall

> THE ONE THING TO KEEP
>
> Study under the conditions of the test: silent, timed, blank paper, nothing to ask.

The mismatch

The exam hall has no chatbot. It has a desk, a clock, and paper.

If every hour of your revision happened with a tool open in the next tab, you have practised a task that does not resemble the one being assessed. Not because you cheated — because the conditions were different, and conditions are most of what determines whether something is retrievable.

Two different rooms

Where the revision happened	Where the assessment happens
• A tab open with something that answers	• A desk, a clock and paper
• Typed, with a delete key and a second attempt	• Handwritten, and hands cramp at forty minutes
• Untimed, and paused whenever it got hard	• Timed, and the reading time is inside the clock
• Warm, because you had just read about the topic	• Cold: you sit down and begin
• Music, a chair you like, a phone within reach	• Coughing, chairs, and nothing at all to consult

This is not a lecture about cheating. Memory is sensitive to the conditions it was formed in, so revision that never resembles the assessment has practised a different task. Closing each of these gaps costs nothing: a kitchen timer, a stack of blank paper, a silent study room and a phone left in another room.

This lesson is about closing that gap deliberately.

The core session

One activity outperforms everything else in the final month, and it is boring.

Sit a past paper, closed-book, timed, phone in another room. No lookups until the clock stops. Write by hand if the exam is handwritten — the motor task matters more than people expect, and so does the fact that you cannot delete a sentence and restart it.

Then, and only then, the AI has a job:

```
Here is the mark scheme and here is my answer. Mark it
strictly. Tell me which marks I lost and why. Do not
rewrite my answer or show me a better one.
```

That final clause is load-bearing. The instant it produces a model answer, you stop diagnosing and start reading, and reading a good answer feels like progress while producing very little of it.

A second prompt worth having:

```
Based on what I got wrong, what do I not understand?
Give me the underlying gap, not the individual errors.
```

Three lost marks scattered across a paper are usually one misunderstanding wearing three costumes. Finding the costume is easy. Finding the misunderstanding is the job.

Space it and mix it

Two findings, both old, both reliable, both ignored every year.

Spacing. Five ninety-minute sessions across three weeks beats one fifteen-hour weekend, and it is not close. A workable schedule for any topic: day 1, day 3, day 7, day 14, day 28. Each session is short, cold, and starts with recall before it starts with notes.

Interleaving. Do not block. Three hours on integration by parts feels efficient and produces a student who can integrate by parts *when told to*. The exam does not tell you. Mixing problem types — so that each question also requires deciding what kind of question it is — feels worse, scores worse in the session, and transfers substantially better.

AI helps here in one specific way: ask it to build you a mixed problem set drawn across four topics in random order, with no labels. Then close the chat and work through it on paper.

The lookup log

Starting today, keep one list: **everything you reached for a tool to get.**

Every formula you couldn't recall. Every date. Every step you needed a hint for. One line each, no self-criticism.

At revision time, that list is your syllabus. It is a far better guide than any feeling you have about your weak topics, because feelings about weak topics are generated by the same fluency illusion that got you here. The log was recorded at the exact moment of not knowing, which makes it honest.

What to stop doing

Rereading notes. Highlighting. And the new one:

Having AI condense your notes into a clean summary and studying the summary. This is now the single most common revision activity among students who have the tools, and it is close to the worst thing you can do with the time. You get someone else's compression, of material you needed to compress yourself, presented with maximum clarity — which produces the maximum feeling of understanding for the minimum retrieval.

If you want the summary, write it yourself from memory, *then* ask the model what you left out. That sequence is one of the best exercises in this course. The other is a summary you cannot make.

The test to apply

> *If you cannot write it on blank paper, in a silent room, with nothing to consult, you cannot write it in the hall.*

Run that check three weeks out, not three days out. Three weeks is enough time to do something about the answer.

One exception

Some courses now assess with AI permitted — open-tool exams, take-home analyses, coursework where the tools are expected. If that is your assessment, the skill being tested is different: it is judgment, verification, and knowing what to ask. Prepare for *that* by practising with the tools under time pressure, and by practising catching their errors.

But find out which kind you are sitting. Do not assume.

BEFORE YOU MOVE ON

Three weeks before the exam, which revision block is doing the most for performance in the hall?

A. Working past-paper questions with the AI alongside, asking whenever you stall, so you get through many more questions.
B. Sitting a past paper closed-book and timed, then having AI mark it against the scheme without rewriting your answers.
C. Having AI condense forty pages of notes into six clear pages, then studying those until everything makes sense.
D. Having AI generate a hundred new practice questions and reading through them with model answers to see the range of what might come up.

66 · 9 min

The Honest Case For Still Learning It

THE ONE THING TO KEEP

You cannot supervise what you cannot do, and supervision is the job that is left.

The question deserves a real answer

"Why learn to do something a machine does instantly and for free?"

That is not a lazy question. It is the right question, and most of the answers students get are bad ones — appeals to character, vague talk about the journey, a teacher who is frightened. Here are the answers that survive scrutiny, in order of strength.

1. You cannot check what you cannot do

These systems are confidently wrong at a rate that matters, and the errors are not random noise. They are plausible. A wrong integral looks like an integral. An invented case citation looks like a citation. A subtly incorrect summary of a paper reads better than the paper.

The only thing between a wrong output and your name on it is whether you know enough to notice. That is a hard limit, and it does not soften as models improve, because as the errors get rarer they also get harder to spot and you get less practised at looking.

This is the argument with no sentimentality in it. Every job that survives the automation of production becomes a job of checking, and checking is a skill made entirely of knowing the thing.

Five answers, strongest first

You cannot check what you cannot do	A wrong integral looks like an integral and an invented case looks like a citation.
Thinking needs material already in your head	Insight is mostly collision, and you cannot collide what you have to look up.
Fluency below frees attention above	Working memory is small. Reconstructing what a derivative is leaves nothing for the model that uses one.
The job argument, stated honestly	Routine production is under real pressure. Nobody knows the size of it.
Some things are worth knowing because knowing them is good	Hearing what a poem is doing. Seeing why a proof has to be true.

The concession belongs beside them, because pretending the line never moves makes the whole case look dishonest: long division of six-digit numbers by hand, log tables, memorising a bibliography format. The test for which side a skill falls on is whether it builds the judgement you will use above it or is purely output. Output can go. Judgement cannot, because there is nothing above judgement to check it.

2. Thinking requires material already in your head

Insight is mostly collision — two things you already hold turning out to be the same thing. You cannot collide what you have to look up, because looking up is serial and slow and you have to already suspect the connection to go looking.

This is why a historian notices something in an economics paper, and why a doctor recognises a pattern in an unusual presentation. Not retrieval speed. Simultaneous possession.

A head with nothing in it, attached to a search engine, is not equivalent. It cannot generate the question.

3. Fluency below frees attention above

Working memory is small — a handful of items. Anything effortful consumes it.

A student who has to stop and think about what a derivative *is* has no capacity left for the model that uses one. A student who has to reconstruct basic grammar has none left for the argument. Automaticity at the lower level is not a nostalgic virtue; it is the precondition for operating at the higher one.

This also tells you which things are genuinely safe to hand over. Anything that never sat underneath something else you do.

4. The job argument, stated honestly

Entry-level work consisting purely of producing routine output — first-pass documents, standard analyses, boilerplate code — is under real pressure. That is not marketing; it is visible in hiring in several fields already.

What nobody knows is the size or shape of it. Anyone giving you a confident number about jobs in 2035 is guessing, including the people selling the tools and the people warning about them.

What you can say is this: "the machine does it, so I don't need to know it" is a bet that *supervision* also gets automated. That may happen. It has not, and every deployment so far has increased the value of someone who can tell good output from bad. Betting the other way is not a hedge. It is a leveraged position on a specific future.

5. The reason nobody says out loud

Some things are worth knowing because knowing them is good.

Being able to read a poem and hear what it is doing. Following a proof and seeing why it has to be true. Holding a century of history in your head so the news makes a different kind of sense. Understanding what is happening in your own body when you are ill.

These change what the world looks like from the inside. That is not an economic argument, it does not appear on a graduate outcomes report, and it does not need to. It is allowed to be a reason.

What is genuinely not worth learning any more

And now the concession, because pretending the line never moves makes the whole case look dishonest.

Long division of six-digit numbers by hand. Log tables. Memorising a bibliography format instead of understanding what a citation must contain. Hand-drawing a graph that any tool draws better. Every generation of teachers has defended a few of these past the point of sense, and students noticed, and it cost the teachers credibility on the things that actually mattered.

The test for which side a skill falls on:

> ***Does this build the judgment I will use above it, or is it purely output?***

Output can go. Judgment cannot, because there is nothing above judgment to check it.

The question to carry out of this course

Not "can a machine do this." Increasingly the answer is yes, and it tells you nothing.

Ask instead:

> ***If the machine does this for me, what will I no longer be able to do?***

Sometimes the honest answer is "nothing I need." Hand it over and use the hour well.

Sometimes the answer is "check its work." Then you already know.

BEFORE YOU MOVE ON

Someone argues: calculators removed the need for mental arithmetic, so AI removes the need to write essays. Where does the analogy break?

A. It does not really break. Writing will follow arithmetic, and courses will adapt to the tool the way they adapted to calculators.

B. It breaks because calculators are reliably correct and AI is not, so the whole difference is one of accuracy.

C. It breaks because writing is self-expression and arithmetic is not, so essays carry a personal value calculators never removed.

D. Arithmetic was output the calculator replaced; the essay is the process that builds the judgment you would need to check any output at all.

---

67 · 9 min

Past papers and mark schemes

THE ONE THING TO KEEP

Past papers are study material rather than a final test, so start in week three, mark yourself against the scheme first, and only then ask which marks were lost and where — never for a model answer.

The highest-yield material you have

If you do one thing differently after this course, make it this: past papers, done under exam conditions, marked against the mark scheme, starting far earlier than feels sensible.

Most students save past papers for the final fortnight, as a test of revision that is already complete. That is backwards. Past papers are the best study material available, not the assessment of it, because they tell you what is actually asked, in what form, with what weighting, and where the marks are.

Start in week three. Do one question badly. That failure tells you what to revise, and the information arrives while you can still act on it.

Reading a mark scheme properly

Mark schemes are published for most public examinations and available from many university departments on request. They are more informative than any revision guide, and almost nobody reads them.

What to look for:

- **What earns the marks.** Frequently a specific term, a specific comparison, or the statement of an assumption — not general quality.
- **How many marks for what.** A four-mark question wants four things. Students routinely write a paragraph for three marks and one line for six.
- **The band descriptors.** The difference between the top two bands is usually one identifiable thing: evaluation, engagement with counterargument, precision of terminology. Find out which, for your subject.
- **The command words.** *Describe*, *explain*, *evaluate*, *justify*, *compare* are technical terms with fixed meanings in most examination systems, and answering the wrong one costs the entire question.

Where the model belongs, exactly

Here is the pattern that works, and the order is the whole of it:

1. **Do the question yourself**, timed, closed book, on paper.
2. **Mark it yourself** against the mark scheme first. Be harsh.
3. **Then** paste the question, the mark scheme, and your answer, and ask: *which marks did I miss, and quote the sentence where each was lost. Do not write a model answer.*
4. Rewrite the parts you lost marks on, yourself.

Step three is legitimate everywhere, produces no submitted text, and is extremely useful. Steps one and two are non-negotiable, because an answer you did not write cannot tell you anything about what you can do.

The request to avoid is *write me a model answer to this question*. You will read it, it will look right, and you will learn nothing — the fluency trap in its purest form.

Where mark schemes do not exist

Many university modules do not publish them. You are not stuck:

- **The learning outcomes** in the module handbook are the criteria in another form. Paste those.
- **Ask the lecturer** what distinguishes a first from a two-one on this paper. Most will tell you, in detail, and are pleased to be asked.
- **Old feedback.** Comments on your returned work are a personalised mark scheme, and the repeated ones are the ones to act on.
- **Ask the model** to construct plausible criteria from the learning outcomes, then check them against a lecturer. Better than nothing; not a substitute for the real thing.

Generating questions when you run out

When you have exhausted the real papers, a model will generate more in the same style, and this is a good use — provided you paste real past questions as examples so it matches the form, and provided you check the generated questions against the syllabus. Generated questions drift towards the plausible and occasionally test something the course does not cover.

A useful prompt:

> *Here are six real exam questions from this paper. Generate eight more in exactly this style, at the same difficulty, covering these syllabus topics. Do not provide answers.*

Do not provide answers keeps them usable as practice rather than reading.

The pattern-finding pass

One genuinely good analytical use. Paste the last ten years of questions and ask which topics recur, in what combinations, and how the phrasing has changed over time.

This is analysis of supplied text, which is the safe category, and it often reveals structure nobody told you about: a topic that appears every second year, two topics that always appear together, a command word that changed in 2022. Check the output against the papers, because it will occasionally over-claim a pattern from thin evidence — which is worth noticing, since inventing patterns from small samples is a failure you should be alert to in your own thinking too.

BEFORE YOU MOVE ON

Why is asking for a model answer to a past paper question worse than marking your own attempt against the scheme?

A. Because reading a good answer produces the feeling of competence without the attempt that would create it.

B. Because a generated model answer will usually contain errors that a real mark scheme would not.

C. Because model answers are written to a generic standard rather than to your examination board's specific criteria.

D. Because using a model answer to revise risks reproducing its phrasing in the exam, which could be treated as an integrity matter.

68 · 8 min

Practising under the conditions of the test

THE ONE THING TO KEEP

Practise in the conditions of the test — timed, silent, handwritten, cold, with nothing to look up — because memory is context-sensitive and coursework-heavy courses hide the gap until the day.

Transfer is narrower than people expect

Memory is sensitive to the conditions in which it was formed. Material learned in one context is retrieved more easily in that context and less easily elsewhere — a well-established finding, sometimes demonstrated in unusually memorable ways, such as divers who learned word lists underwater recalling them better underwater than on land.

The practical consequence for you: if every hour of revision happens in a comfortable chair with music, a window open and a phone in reach, and the exam happens in silence, timed, by hand, with nothing available, you have practised a different task from the one you will perform.

The conditions that actually differ

Make a list for your specific assessment and close each gap deliberately.

- **Time.** Almost everyone underestimates how much slower they are under pressure. Practise at the real pace, including reading time.
- **Handwriting.** If your exam is written by hand and your practice is typed, you have not practised the physical act, and hands cramp at forty minutes. Write by hand, at length, before the day.
- **No lookup.** Any practice where you check something mid-answer is not practice for a closed exam. Put the answer you are unsure of in and carry on; mark it afterwards.

- **Silence, or noise you cannot choose.** Exam halls have coughing, chairs, a clock. Some people find a library reading room a good approximation.
- **Cold start.** In an exam you sit down and begin, without a warm-up on the topic. Practise starting cold.
- **Question order and choice.** If your paper offers choice, practise choosing quickly. Time spent deciding is time not writing, and students lose a surprising amount to it.

The exam that allows AI

Some assessments now permit these tools, and open-book, open-AI examinations are appearing. They are not easier. When everybody has the tool, the questions change: they ask for judgement, for application to a novel case, for critique of a supplied answer, for defence of a choice.

If your assessment permits it, practise with it under time pressure, because using a model well under a clock is its own skill. The specific failure to rehearse against is spending eight minutes prompting your way around a question you could have answered in three. In a timed assessment the tool is often slower than knowing the answer, and knowing which is which is exactly what is being tested.

Coursework-heavy courses have a hidden risk

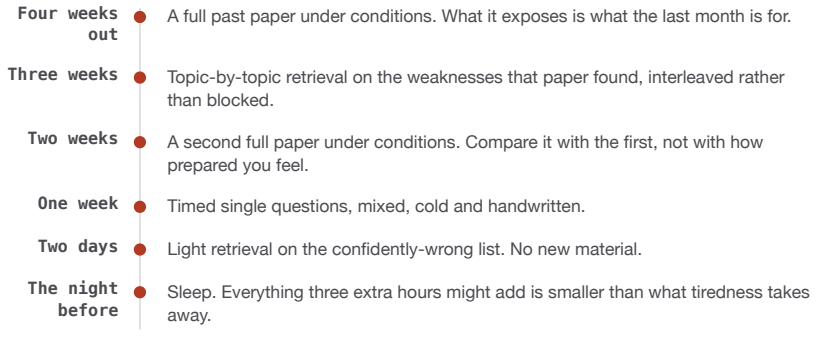
If your degree is assessed mostly by coursework, with one closed component, you can arrive at that component with excellent marks and no unaided capability. The gap is invisible until the day.

The protection is the unaided check from Module 3, run regularly, all year. It costs an hour a month and it is the difference between finding out in week five and finding out in the hall.

Building the exam week backwards

A schedule that works, counting back from the date:

Counting backwards from the date



Each session is the same idea at a different distance: practise the task you will be assessed on, in the conditions you will be assessed in. A course assessed mostly by coursework hides the gap until the day, which is why the unaided check belongs in week five rather than in the hall. Everything here is free — a timer, blank paper and a room you are not comfortable in.

- **Four weeks out.** Full past paper under conditions. Mark it. This tells you what the last month is for.
- **Three weeks.** Topic-by-topic retrieval on the weaknesses that paper exposed. Interleaved.
- **Two weeks.** Second full paper under conditions. Compare.
- **One week.** Timed single questions, mixed, cold. Handwritten.
- **Two days.** Light retrieval on your confidently-wrong list. No new material.
- **The night before.** Sleep. This is not a platitude — sleep consolidates memory, and a night lost to cramming costs more than the material gained. Everything the extra three hours might add is smaller than what tiredness takes away.

The free way to simulate the hall

You do not need anything bought. A kitchen timer or a phone alarm, a stack of blank paper, a pen, and a room you are not comfortable in. Your university library usually has silent study rooms that are much closer to exam conditions than your bedroom, and they cost nothing.

A group of three doing a past paper together in silence, in the same room, at the same time, then swapping and marking each other's against the scheme, is

the closest free approximation to the real thing, and the marking-somebody-else's half is where a surprising amount of the learning happens.

On the day

Read the whole paper first. Allocate time by marks, and write the allocation on the paper. Start with a question you can do, to get moving. Leave the last ten minutes to check.

And if a topic you revised does not appear, that is normal, and it is not evidence you revised wrongly. Papers vary. The students who do well are usually not the ones who predicted the paper; they are the ones who could answer whatever was on it, which is what interleaved practice buys you.

BEFORE YOU MOVE ON

Why are open-AI examinations generally not easier than closed ones?

A. Because the time pressure is greater, since prompting and reading the output consumes minutes that writing would have used.

B. Because candidates who rely on the tool produce generic answers, and generic answers score in the middle band regardless of the question asked.

C. Because access to a model is unreliable under exam conditions, so candidates cannot depend on it.

D. Because the questions change to ask for judgement, application to novel cases and defence of choices.

69 · 8 min

Vivas, and being asked about your own work

THE ONE THING TO KEEP

A viva tests why you made each choice and what you rejected, so rehearse out loud against a model playing a difficult examiner — and a calm "I do not know, but I would look at X" scores better than a confident invention.

Assessment is moving towards the spoken

As written work becomes harder to attribute, institutions are shifting weight towards forms where you are present: vivas, oral defences, in-class written tasks, presentations with questions, and short interviews about submitted coursework. Several universities now run a viva as a routine follow-up to any coursework that raises a question, and some have made short oral components standard.

This is not primarily a policing measure, though it functions as one. Being able to explain your own work under questioning is a genuine professional skill, and it is the one form of assessment that has never been vulnerable to any of this.

What is actually being tested

Not recall of your document. Whether you can:

- Explain **why** you made each choice — this method, this structure, this source.
- Say what you **rejected** and why. This is the strongest signal of ownership and the hardest thing to fake.
- Identify your own work's **weaknesses** before they are pointed out.
- Answer a question you have not anticipated by reasoning aloud from what you know.

- Say **I do not know** when you do not, and then say how you would find out.

That last one is worth internalising. Examiners are not trying to find something you cannot answer as a trap; they are finding the edge of your knowledge, which is their job. A calm *I do not know, but I would look at X* scores better than a confident invention, and inventing under questioning is transparent to anybody who knows the field.

Preparing, with the tool that is genuinely excellent here

This is the single best use of a language model in assessment preparation, and it is permitted everywhere because it produces nothing you submit.

> *Here is my dissertation. Act as an examiner in this field. Ask me ten questions, one at a time, starting with the obvious and moving to the hardest. After each answer, tell me whether it was adequate and what an examiner would follow up with. Be difficult.*

Then answer **out loud**. Not in your head, and not typed. The gap between knowing something and being able to say it under mild pressure is exactly what a viva measures, and the only way to close it is to have said it before.

Do this three times across a week and the actual event becomes recognisable rather than novel.

The questions that come up almost always

Whatever the field, prepare these:

- **Summarise your work in two minutes.** Practise until it is fluent. It is the opening question in most vivas.
- **What is your contribution?** What is here that was not here before.
- **Why this method and not the obvious alternative?** Know the alternative and why you rejected it.
- **What is the weakest part?** Have an honest answer. Claiming there is none is the worst possible response.

- **What would you do differently?** Shows judgement developed through doing the work.
- **What would you do next?**
- **Explain this figure / this paragraph / this line of code.** Chosen at random. Know your own document.

If you used AI, be ready to say how

In a viva about coursework, expect to be asked. Answer plainly, matching your disclosure statement. *I used it to critique my structure and to explain two papers I could not follow; the argument and the writing are mine* is a complete answer, and it is far better than a vague one.

Be ready to demonstrate it, too: explain the argument, discuss the sources, describe the changes you made and why. A student who used the tool as this course describes will find this easy. That ease is the point of the whole course.

Nerves, practically

Everybody is nervous and examiners expect it. Three things help more than reassurance: having said the answers out loud beforehand, having water and a copy of your own work in front of you, and having permission in your own head to pause. A three-second silence before answering feels enormous to you and is unremarkable to everyone else.

And if you go blank, say so and ask to come back to it. That is normal, it happens in most vivas, and it costs nothing.

BEFORE YOU MOVE ON

Why is being able to say what you rejected the strongest signal of ownership in a viva?

A. Because rejected options are not in the submitted document, so they cannot have been supplied by any tool.
B. Because the reasons for rejecting an option exist only in the process of doing the work.
C. Because examiners use rejected alternatives to assess breadth of reading rather than depth of understanding.
D. Because a viva is scored partly on methodological awareness, and naming alternatives demonstrates that criterion directly.

70 · 9 min

Long projects and dissertations

THE ONE THING TO KEEP

Write a working document from week one — it converts reading into thinking, removes the blank page, and produces the strongest possible process record — and keep the dated file of how your question changed.

A different kind of work

A dissertation or final-year project fails in ways a three-week essay does not. Not usually through inability, but through drift: months in which the question changes shape, the reading expands without converging, and nothing is written until the last six weeks.

What follows applies to any long piece — a research project, a capstone, a portfolio, a design project.

The question, and its instability

Your question will change. That is normal and it should be deliberate rather than accidental. Write it down, dated, in one sentence, and rewrite it whenever it moves, keeping the old versions.

That file of successive questions is worth a great deal at the end: it is the intellectual history of the project, it makes writing the introduction straightforward, and in a viva it demonstrates exactly the judgement you are being examined on.

A good use of a model early: *here is my question. What would I need to establish to answer it? What is ambiguous in how I have phrased it? What would somebody object to about the scope?* Diagnosis of your question, not a supply of one.

Write from week one

The single most useful habit. Not the dissertation — a working document, updated weekly, containing what you have read, what you now think, and what you do not know.

This does three things. It converts reading into thinking as you go, rather than in a panic at the end. It means you always have text, so the blank page never arrives. And it produces a continuous, timestamped record of the work developing, which is the process record from Module 7 in its strongest form.

Students who write only at the end lose most of what they read in months one and two, because they never encoded it.

The literature, without drowning

A method that converges:

1. **Start with two or three recent reviews.** They give you the map and a curated reference list.
2. **Follow citations backwards** to the foundational work everybody cites, and **forwards** through Scholar's *cited by* to the recent work.

3. **Stop when you stop meeting new names.** Convergence is the signal, and it arrives sooner than you fear.
4. **Four lines per paper**, in your own words, in Zotero, from Module 5.

A model is useful for step zero — the vocabulary a field uses, so your searches match — and for explaining methods sections you cannot follow. It is not useful for producing the reference list, and a bibliography you did not build is a bibliography you cannot defend in a viva.

Supervision

Your supervisor is the most valuable resource in the project and the most under-used. Practical points:

- **Send something before every meeting**, even if it is bad. A meeting about a document is worth five about intentions.
- **Ask specific questions.** *Is my second research question answerable with this data* gets a useful answer; *how am I doing* does not.
- **Write up every meeting** — what was agreed, what you will do — and email it to them. This prevents the common disaster of two people remembering different agreements, and it is a record.
- **Say when you are stuck**, early. Supervisors have seen every version of stuck and mostly want to help. The failure mode is disappearing for six weeks out of embarrassment.

Ethics and data

If your project involves people, ethics approval comes before data collection, not after, and there is no retrospective route. Your approval says where data will be stored and who may access it — a commercial AI service is almost certainly not on that list, and uploading participant data is a research-integrity breach independent of any AI policy. Anonymise before anything leaves your approved storage, remembering from Module 2 that removing names is not anonymisation when the combination of details identifies somebody.

Where AI helps across the project

Honestly and specifically: understanding difficult papers, explaining statistical methods, generating viva questions, critiquing your structure, checking your draft against the criteria, debugging analysis code, and helping you say clearly something you already decided to say.

And where it does not: choosing the question, deciding what the data means, judging which literature matters, or writing anything you would be asked to defend. Those are the parts being assessed and the parts you will be examined on.

The schedule that works

Count backwards from submission, and put the last month aside for writing and revision, not for new work. Then:

- **Reading and question refinement** — first third.
- **Data collection or core work** — second third, and it will take longer than planned.
- **Analysis and writing** — final third, with the last month for revision only.

Everybody overruns the middle. The way to protect the end is to start writing in the first third, so the final month is genuinely revision rather than composition.

BEFORE YOU MOVE ON

Why does keeping every dated version of your research question turn out to be so useful at the end?

A. Because it demonstrates that the project was completed within the approved scope, which is checked at submission.
B. Because supervisors expect to see how the question developed and mark the project partly on that development.
C. Because it records the reasoning behind each change, which is what an introduction and a viva both require.
D. Because a question that changed repeatedly indicates genuine engagement, which distinguishes an original project from a derivative one.

71 · 8 min

Applications, statements and interviews

THE ONE THING TO KEEP

The test for every sentence in an application is whether another applicant could have written it — so use the model to ask you questions that surface specific stories, never to produce the statement.

Why the generated personal statement fails

Admissions tutors and recruiters read thousands of these. A generated statement is recognisable not because of any detector, but because it says the same things: passion for the subject from a young age, a formative moment described in general terms, transferable skills, a closing sentence about contributing to the community.

Every one of those is the average of what has been written, which is precisely what a next-token predictor produces. The average is exactly what does not distinguish you, in a document whose only purpose is to distinguish you.

And increasingly it is checked. Several admissions systems and employers now ask questions at interview that only somebody who wrote the statement can answer, and some application processes have moved to short live tasks for this reason.

What actually works

Specificity, and it is the whole answer.

Not *I am fascinated by biology* but *I spent a summer counting bees in my grandmother's garden and found that the count changed with the time of day more than with the weather, which is when I started reading about foraging behaviour.*

The second is memorable, unfakeable, and true. It also tells the reader far more: that you notice things, follow them up, and read beyond the syllabus. None of that had to be asserted.

The test for any sentence in a personal statement: **could another applicant have written this?** If yes, cut it. What survives is short, and it is yours.

The legitimate uses

There are real ones, and they are all at the edges:

- **Interrogation.** *Ask me fifteen questions about my experience that would surface specific stories.* Excellent. Most people cannot remember their own material on demand, and questions unlock it.
- **Structural feedback on your draft.** Which paragraph is vague, where you asserted instead of showing.
- **Checking against the criteria**, where an application publishes them.
- **Grammar and spelling.** Ordinary.
- **Interview preparation** — generating likely questions and practising aloud. Very good, and the same technique as the viva lesson.

The rules, which are real

Some application systems now explicitly prohibit AI-generated statements and require a declaration. UCAS in the UK, for example, expects the statement to be the applicant's own work and screens for similarity; universities and employers have their own positions. Fabricating experience is fraud, and it has consequences beyond a rejection — offers are withdrawn and, where a qualification results, revoked.

Read what the application says and follow it. If it asks whether you used AI, answer honestly; a truthful yes about grammar checking has never cost anybody a place.

CVs and cover letters

A slightly different case. These are formulaic documents, the form is genuinely standardised, and using a tool to structure them is normal practice — many employers use automated screening at the other end of the same pipe.

What still has to be yours: every claim of fact, every number, and every example. A generated cover letter that describes achievements you did not have is a lie you will be asked about in an interview, by somebody holding it.

The good use is the reverse of generation: paste the job description and your real experience, and ask which of your experiences match which requirements, and what you have not evidenced. Analysis of supplied material, and it is genuinely useful, because most people are bad at seeing which of their own experiences are relevant.

Interviews

The part no tool sits in, and where everything above gets tested.

- Practise out loud, against generated questions. The same technique that works for a viva.
- Have three or four specific stories ready that you can adapt to different questions. Real ones, with details you remember because they happened.

- Prepare to be asked about anything on your application, including the thing you mentioned in passing.
- Have a question of your own that shows you looked into them properly.

And the connection back to this whole course: an interview is an unaided check. Everything you can actually do is available to you there, and everything you cannot is not. Which is why the habits in Module 3 turn out to matter for reasons well beyond exams.

BEFORE YOU MOVE ON

Why does a generated personal statement fail even when it is well written and contains no errors?

A. Because admissions systems screen statements for AI generation and reject flagged applications.

B. Because the register is too polished for an applicant of that age, and the mismatch between the statement and the rest of the application raises doubt.

C. Because it cannot include specific personal experiences, which are what admissions tutors are looking for.

D. Because it produces the average of what has been written, in a document whose only purpose is to distinguish you.

72 · 8 min

Placements, and rules that are not your university's

THE ONE THING TO KEEP

On placement three rulebooks apply at once and the strictest governs — the organisation's information does not leave its systems, and a professional body's fitness-to-practise process is a far heavier matter than academic misconduct.

Three rulebooks at once

On a placement — clinical, teaching, legal, engineering, or an ordinary internship — you are simultaneously governed by your university's academic policy, the host organisation's own rules, and often a professional body's code of conduct. They do not say the same things, and the strictest one applies.

Students get into difficulty here not through dishonesty but through applying a university rule in a setting where a different and stricter one governs. It is worth ten minutes on day one to find out which rules you are actually under.

The host organisation's rules come first

Ask, in your first week, and ask a person rather than assuming:

- Is there a policy on using AI tools with organisation data?
- Are there approved tools, and is anything prohibited?
- What may leave the organisation's systems, and what may not?
- Who do I ask when I am not sure?

Many organisations now have a written policy; many do not and are working it out. Either way, the answer you get in writing is your protection. Where there is genuinely no policy, the safe default is that nothing belonging to the organisation goes into any external tool.

Confidentiality is not an AI question

The rule that matters most on placement predates all of this. Patient information, client information, pupil information, commercial information: it does not leave the systems it is meant to live in. Pasting it into a chat window is a disclosure to a third party, and it is treated as one.

This is the Module 2 rule with real consequences attached. A nursing or medical student who puts a case into a general assistant has made a data breach, regardless of intent, regardless of whether names were removed, and regardless of what the university's AI policy says. Teaching placements are the same for pupil information; legal placements the same for client matters. Professional bodies discipline for this, and a fitness-to-practise process is a much heavier thing than an academic misconduct process.

The workable version: discuss the *general question* — the mechanism, the guideline, the legal principle — without the particulars. That is a legitimate use and it is what a textbook would give you anyway.

Reflective writing, which is a special case

Many placements require reflective accounts, and these are the least suitable thing in the world to generate. A reflective piece is assessed on whether you noticed something and thought about it. A generated reflection is a description of what reflection sounds like, and assessors in these professions read hundreds of them a year and can tell.

It is also self-defeating: reflection is one of the few written exercises whose entire benefit is in the doing. Write it badly and truthfully rather than well and generically. *I did not know what to say to the patient and I said something clumsy* is worth more than any amount of polished insight.

The safe assistance: write the reflection yourself, then ask for structural feedback, or ask questions that prompt you to go deeper. Anonymise anything you paste, and check whether your placement permits pasting it at all.

Professional codes

If you are heading into a regulated profession — medicine, nursing, teaching, law, engineering, accountancy, architecture, social work — the regulator has a code, it applies to students in many cases, and it usually says something about honesty, confidentiality and competence. Read yours once. It is short, it is free, and it is the standard you will be held to for the rest of your career.

The recurring theme across all of them: you are accountable for work issued in your name, and you cannot delegate that accountability to a tool. That principle resolves most questions before they arise.

Using it well on placement

There is a great deal that is entirely fine and genuinely useful:

- Understanding a guideline, a procedure, or a piece of legislation in general terms.
- Preparing for a difficult conversation by rehearsing it.
- Explaining a concept you met on the ward, in the classroom, or in the office and did not follow.
- Drafting your own learning objectives, or structuring your portfolio.
- Practising the questions your supervisor might ask.

All of these are about you and your learning, and none of them involves the organisation's information leaving its systems.

When you are unsure in the moment

Do not resolve it yourself under time pressure. Ask your practice supervisor or your placement lead — that is what they are for, and asking is treated as good practice rather than as ignorance. The students who get into trouble are almost never the ones who asked.

BEFORE YOU MOVE ON

A student on a clinical placement checks their university's AI policy, confirms that AI use with disclosure is permitted, and pastes an anonymised case into an assistant. What have they got wrong?

A. The host organisation and the professional code govern, and both are stricter.

B. Anonymisation is never sufficient for clinical material, so no discussion of a case is permissible in any form.

C. The university policy requires disclosure at the point of use rather than at submission, so the declaration came too late.

D. Assistants retain conversation data, so anonymised clinical material could be recovered later from the provider's stored logs.

73 · 9 min

What to learn anyway

THE ONE THING TO KEEP

The shift is from production towards judgement, and judgement is built by production — you cannot supervise a skill you never acquired, which is why the argument for learning survives the tools that do the producing.

The question, asked honestly

If a machine can write competent prose, produce working code, summarise a literature and pass many examinations, what is the point of learning to do those things?

The question deserves better than reassurance. Some of what was taught for decades was valuable mainly because it was scarce, and some of that scarcity has gone. Pretending otherwise insults the person asking.

Here is the honest answer, in parts.

You cannot supervise what you cannot do

The jobs that remain in every field where these tools are used well are jobs of judgement: deciding what to ask for, evaluating what comes back, and being accountable for it. All three require the underlying competence.

A doctor who cannot read an ECG cannot tell when the model's reading is wrong. An engineer who cannot do the calculation cannot tell that the output is off by a factor of ten. A lawyer who does not know the area cannot tell that the case is invented — and several have found out expensively. The people who use these tools most effectively are not the ones who know least; they are the ones who know enough to check.

This is the strongest argument and it is entirely practical. Supervision is the work that is left, and supervision is unavailable to somebody who never acquired the skill.

Understanding is not a means to an end

A weaker-sounding argument that is nonetheless true. The point of knowing how something works is not only to produce outputs. It is to be able to think about it — to notice that two things are connected, to recognise a pattern from another field, to be interested.

Somebody who has understood how evolution works sees the world differently, and that does not go away when a machine can also describe evolution. The value was never in being the only one who could produce the description.

What has genuinely changed

Not everything is unaffected, and it is worth saying which things.

- **Producing competent first-draft prose** is less scarce than it was.
- **Boilerplate code** is much less scarce.
- **Translation** is far cheaper.
- **Summarising a document** is nearly free.

What has become more valuable, correspondingly:

- **Judging quality**, which requires knowing the field.
- **Asking the right question**, which is most of research and most of engineering.
- **Being accountable**, which is a human position that cannot be delegated to a system nobody can question.
- **Doing the physical and interpersonal parts** of almost every job.
- **Knowing what is true**, and being able to check.

The shift is from production towards judgement. Judgement is built by production. That is the whole of the answer, and it is why the argument for learning survives.

What to deliberately keep

Revisit the list from Module 1, now with a term or a year of evidence:

1. The five or six capabilities your field is built on. For you, specifically.
2. The ability to read something difficult, slowly, without help.
3. The ability to write something from a blank page.
4. Enough arithmetic and estimation to notice when a number is wrong.
5. The ability to be wrong in front of somebody and keep going.

That last is not a joke. Everything in this course that works — the blank page, the unaided check, the viva rehearsal, the group argument — requires tolerance for being visibly not-yet-good at something. The tools remove the discomfort, and the discomfort was doing work.

The thing nobody can tell you

What the next ten years will require is genuinely unknown. Confident predictions in both directions are being made by people with no better information than you have. Anybody claiming to know which skills will be obsolete is guessing, including the people selling courses about it.

What is defensible: capabilities that let you learn new things quickly, judge unfamiliar material, and be trusted by other people have been valuable across

every previous technological change. That is not a guarantee. It is the best available bet, and it happens to be exactly what an education, done properly, produces.

The end of the course

Everything here reduces to one habit, and it is the one to leave with.

After each session with these tools, ask what changed: the document, or you. Both are legitimate answers on different days. But if the answer is *the document* every time, for a year, then at the end of that year you will have a great many documents and you will not be able to do anything you could not do at the start.

That is the whole risk, and it is entirely within your control.

BEFORE YOU MOVE ON

What is the central practical argument in this lesson for still acquiring a skill a machine can perform?

A. That the tools will not be reliably available, so the underlying capability is needed as a fallback.

B. That judging and being accountable for the output requires the competence itself.

C. That employers continue to test these skills at interview, so they remain necessary for getting hired.

D. That skills learned in one domain transfer to others, so acquiring any difficult capability improves general reasoning ability.

CASE STUDY · MODULE 8

Ten minutes each, two hundred and forty students

A computer applications department in Madurai deciding whether to put ten marks of a coursework module onto a viva nobody had the hours to run.

The second-year programming module had been assessed entirely by coursework since 2021: four assignments, no closed component, all submitted from wherever the student happened to be. Marks had risen every year since. The staff were not, on the whole, suspicious of the marks, and nobody in the department wanted a conversation about cheating. They were worried about something narrower and more concrete.

Two employers who had recruited from the department for years had stopped coming. Both said a version of the same thing when asked. Candidates arrived with strong transcripts, could not write a small function on a whiteboard, and could not say what their own final-year code did when it was put in front of them.

Dr S. Lakshmi, who chairs the department's assessment committee, had three options, and a regulation removed one of them before the discussion started.

Reinstating a closed written examination meant changing the module's assessment weighting, and the university's regulations require a full session's notice for that. It was the right answer for the year after next and no help whatever in the year they were in.

Leaving it alone was defensible on the numbers and indefensible on the thing the numbers were supposed to stand for. It also, she pointed out, protected nobody: a student whose transcript says eighty-two and who cannot write a function is going to find that out in an interview room instead, in front of somebody with no reason to be kind about it.

The third option was to move ten of the hundred marks onto a ten-minute viva on the student's own submitted code. Why this data structure and not the obvious alternative. What did you try first and abandon. Explain line forty.

The first cost was arithmetic and it was not negotiable. Two hundred and forty students at ten minutes each is forty hours, before scheduling, before the two minutes of turnover between students, in a week when the same five staff are marking everything else. There was no money for extra hours and no realistic prospect of any.

The second cost is the one the committee argued about for an hour, and it is the serious one. A viva measures speaking under mild pressure alongside the thing it is meant to measure. It falls hardest on students who are shy, on students whose spoken English is weaker than their written English, and on anybody who goes blank when a clock is running. None of those is what a programming module is supposed to assess. A student who wrote every line of their own work and cannot say so in ten minutes loses marks for a reason that has nothing to do with programming, and the department would never know it had happened.

They went ahead, with three adjustments aimed squarely at that second cost.

Twenty questions were published a week in advance and three drawn at random on the day, so the form could be rehearsed even though the content could not be memorised. Every student could bring a printed copy of their own code into the room and refer to it throughout. And a written alternative was arranged through the disability service for anybody with a documented adjustment, set up in advance, requiring no explanation to the examiner sitting opposite them.

STOP HERE

Before turning the page: what would you have done, and what would it have cost if you were wrong? Write it down. The point of the next section is to compare.

WHAT HAPPENED

The vivas took forty-one hours across five staff and one week. The average was 6.2 out of ten.

The relationship with coursework marks was much weaker than anybody had predicted. Of the thirty-one students with coursework above eighty-five, six scored below four. Of the students sitting between fifty-five and sixty-five, nine scored nine or ten.

Karthik had a coursework average of eighty-two and could not say what his own recursive function returned for an empty list. He was not accused of anything, because there was nothing to accuse him of: the ten marks were simply not earned. He sat the viva again in the resit window after a month of ten-minute unaided checks on blank paper, and scored eight.

Fathima had a coursework average of sixty-one and scored ten. Asked what she had tried first, she named two approaches she had abandoned and why. She keeps a plain text file of unaided checks going back to her first term: a date, a task, and whether she finished it without help.

Two staff described the week as the most informative of the year, on the grounds that they now knew what the cohort could actually do rather than what it could submit.

A student representative took it to the staff-student committee and argued that the viva measures confidence rather than competence. The department's answer was the published question list and the written alternative. Her second point was accepted immediately and fixed: the question list had gone out four days late, on the portal only, in English only, which had quietly removed most of the benefit it was supposed to provide.

One of the two employers returned the following year. The other has not.

Six students with coursework above eighty-five scored below four. Give two different explanations, and say how the department could tell them apart.

One is that the coursework was not substantially theirs, in which case there is nothing to recover and the viva has found it. The other is that it was theirs and the capability did not survive the conditions: no editor, no lookups, no second attempt, and a person waiting while you speak. These are separable and the department already has the instrument. A student in the second group can rebuild the capability, which is exactly what Karthik's resit demonstrated after a month of unaided

checks, and can usually describe what they tried and abandoned even when they stumble over the specifics. A student in the first group cannot answer what did you try first, because the question is not about the code, and no amount of preparation from the submitted file supplies an answer to it.

The three adjustments were aimed at the fact that a viva measures speaking too. What did each one do, and what did none of them fix?

The published question list removes surprise about the form, so the thing can be rehearsed out loud in advance, which is the only way the gap between knowing something and saying it under pressure closes. The printed code removes the memory load, so nobody loses marks for not recalling their own line numbers. The written alternative through the disability service makes the adjustment formal and documented, which means it is a complete answer rather than something a student has to explain in the room. None of them removes the fact that an examiner is present and the clock is running, and none of them helps a student who has never said any of it aloud before the day. That last gap is closed by rehearsal and not by policy, which is why practising against a model playing a difficult examiner is worth three evenings.

Why is 'what did you try first' harder to answer from a submitted file than 'what does this line do'?

The file records the destination and not the route. What a line does can be read off the code by anybody competent, including somebody seeing it for the first time, so the question tests reading rather than authorship. What you tried and rejected exists only in the history of the work: the approach that was too slow, the structure that broke on an empty input, the thing you did for two hours before realising it would not generalise. It cannot be reconstructed from the artefact and it is the single hardest thing to fabricate under questioning, which is why examiners ask it and why it is also the strongest evidence of ownership a student can offer. Fathima could answer it because she had a record of her own attempts, kept for her own reasons, long before anybody asked her for one.

MODULE 8 · TEST

Check what stuck

6 questions, one right answer each. This one is for you, not for a record: answers and the reason behind each are at the back. If two questions from the same module go wrong, re-read that module before the exam rather than after.

1 When should past papers enter your revision, and in what role?

A. In week three, as the best study material available
B. In the final fortnight, as a test of completed revision
C. Only once the mark scheme has been read from end to end
D. After a model answer has been generated for comparison

2 Why does it matter that you revise in the conditions of the assessment rather than just on the material?

A. Handwriting looks better to a marker than typed answers do
B. Silence makes the brain consolidate memories more quickly
C. Memory is sensitive to the conditions it was formed in
D. Timed practice raises the number of questions you attempt

3 What is a viva about your own coursework actually testing?

A. Whether you can recall your document accurately from memory
B. Whether you can speak fluently for ten minutes without notes
C. Whether your answers match the examiner's own position
D. Why you made each choice, and what you rejected

4 What is the hidden risk in a degree assessed almost entirely by coursework?

A. Coursework marks are moderated more harshly than exam marks
B. High coursework marks can hide no unaided capability
C. Coursework deadlines cluster, which lowers the average grade
D. Coursework modules carry fewer credits than examined ones

5 What is the test for whether a skill is still worth acquiring yourself?

A. Whether a machine can currently produce it faster than you can

B. Whether it appears in the job advertisements you are reading

C. Whether it builds judgement you will use above it

D. Whether your institution's policy permits assistance with it

6 Your examination permits AI tools. Which failure is worth rehearsing against?

A. Prompting your way around a question you could have answered

B. Being unable to reach the tool because the network is busy

C. Forgetting which model the examination board has approved

D. Running out of free-tier messages before the paper is finished

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, and preparing for assessment that happens with nothing to consult. Twenty-five questions are drawn at random from a larger pool, so no two papers are the same. The pass mark is 70 per cent, and passing opens the next course in the track. There is no timer: many readers here are working in a second or third language, and a clock measures panic alongside understanding. If you do not pass, you can retake after thirty minutes and the questions will be drawn fresh. A teacher can open the next course for you at any time regardless of this paper, and that is recorded as an override rather than disguised as a pass.

On the website a sitting is 25 questions drawn from this pool and the pass mark is 70%. There is no time limit, there and here. Passing opens the next course in the track.

1 `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 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 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 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 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 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 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 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

9 2. Asking well

You want a model to attack an answer it has just produced. What should you do?

A. Ask at the end of the thread that produced it
B. Paste it into a fresh chat and ask for the weakest step
C. Ask it to rate its own answer out of ten first
D. Tell it the answer came from a classmate rather than itself

10 2. Asking well

Which of these symptoms means that no amount of rewording will help?

A. The answer is generic and starts from the beginning
B. It contradicts something it said twenty messages ago
C. It is vague about your course's particular notation
D. You asked it to count the words in your essay

11 2. Asking well

In the quiz-me prompt, which clause is doing the work that makes it retrieval rather than reading?

A. One question at a time, and wait for my answer
B. Make the questions progressively harder as we go
C. Keep going until we have covered the whole chapter
D. Give me the answer with each question for reference

12 2. Asking well

A nursing student wants help thinking about a case from placement at eleven at night. What is the workable version?

A. Discuss the clinical question in general terms
B. Paste it with the patient's name removed
C. Paste it into the institutional account instead
D. Paste it after switching the training toggle off

13 2. Asking well

Why ask for the assumptions as a separate list before the solution?

A. It stops the model using an approximation at all
B. It surfaces the thing you actually need to check
C. It guarantees the chain shown is the one that was used
D. It makes the answer shorter and quicker to read

14 2. Asking well

You set custom instructions saying you are a second-year. What is the standing cost?

A. They are re-sent with every message and use up quota
B. They stop the model from searching the web at all
C. They apply only to the chat in which they were set
D. They will simplify things you later need in full

15 3. Study that raises the difficulty

Which kind of stuck should you resolve immediately, with no learning cost?

A. You have an answer and cannot see why it is wrong
B. You know the method and the execution has gone wrong
C. You cannot remember a formula you once knew
D. You chose the wrong method and have not noticed yet

16 3. Study that raises the difficulty

Under what condition is struggle a learning condition rather than a bad afternoon?

A. When it is followed by instruction
B. When it lasts at least thirty minutes without a break
C. When it happens with no tool available at all
D. When the problem is harder than the assessment will be

17 3. Study that raises the difficulty

Why does writing eight questions on a chapter beat having eight written for you?

A. Generated questions are frequently factually wrong
B. Your questions match the examiner's style more closely
C. It takes less time than answering a generated set
D. Writing one requires knowing the load-bearing ideas

18 3. Study that raises the difficulty

Why quiz yourself from your own pasted notes rather than from the model's general knowledge?

A. Pasted notes cost fewer tokens than a general question
B. You cannot audit errors in material you do not know
C. The model refuses to quiz without a source supplied
D. Notes are more current than the model's training data

19 3. Study that raises the difficulty

After explaining a topic from memory, which feedback request separates understanding from a correctly memorised rule?

A. Tell me which parts of my explanation were vague
B. Tell me what I appear to have memorised without understanding
C. Tell me what I omitted entirely from my account
D. Tell me whether my explanation was clear enough for a beginner

20 3. Study that raises the difficulty

Which of these has been tested repeatedly and does not hold up?

A. Teaching to a claimed learning style
B. Rewriting your notes from memory after a lecture
C. Spacing reviews at increasing intervals
D. Mixing problem types without labelling them

21 3. Study that raises the difficulty

A fifty-minute session collapses in one predictable way. Which?

A. The break is skipped and attention fades early
B. The blank page takes longer than five minutes
C. The assisted block expands and eats the rest
D. The interleaved questions come from one topic

22 4. Subject by subject

Why is a generated quotation from a novel more damaging than a generated summary of it?

A. Quotations are copyright and summaries are not
B. An argument on words not in the book is worth nothing
C. Summaries are checked against more sources in training
D. Markers penalise quotation marks in undergraduate essays

23 4. Subject by subject

When should you consult the critical consensus on a set text?

A. Before reading, so you know what to look for
B. While reading, to check your notes as you go
C. After you notice and write your own observation
D. Only if your tutor names a critic in the seminar

24 4. Subject by subject

What makes an agreeable AI conversation partner risky for a language learner?

A. It uses vocabulary above the level you asked for
B. It understands your errors and continues past them
C. It corrects too aggressively and damages confidence
D. It defaults to a formal register in every language

25 4. Subject by subject

Twelve students in each of four classes are measured, and the forty-eight results are analysed as independent observations. What is wrong?

A. The sample is too small for a parametric test
B. The groups are unequal and need Welch's correction
C. Normality should be tested on the raw data first
D. Observations within a class are clustered

26 4. Subject by subject

Generated essays fail the same way across every essay subject. What is wrong with the thesis?

A. The claim is stated too early in the introduction
B. The claim is contestable but poorly evidenced
C. The claim is one nobody would dispute
D. The claim uses terminology from another discipline

27 4. Subject by subject

Why is a confident statement of the law especially dangerous?

A. Legal language is denser than other academic prose
B. Models blend jurisdictions, and the text does not show it
C. Case law changes faster than any training cut-off allows
D. Statutes are rarely included in any training corpus at all

28 4. Subject by subject

Why do significant figures matter more in a lab report than they look?

A. They show whether you understood your measurement
B. They are required by the journal's submission format
C. They make the calculation quicker to check by hand
D. They determine which statistical test is appropriate

29 5. Research, sources and evidence

A lawyer asked the chatbot whether the cases it had cited were real, and it said yes and produced full text. Why did that not help?

A. The tool's search feature had not been switched on
B. He asked in the same chat that produced them
C. It has no channel to the world
D. He should have asked for the docket numbers as well

30 5. Research, sources and evidence

You have the full text of a paper open. Which step turns the citation into evidence?

A. Saving it into Zotero with the PDF attached
B. Recording the page your claim rests on
C. Checking the journal is in the DOAJ listing
D. Adding it to your reference list in the right style

31 5. Research, sources and evidence

Of the four lines in a note on a paper, which do students most often leave out?

A. What they say they showed
B. The design, n and key number
C. What the paper cannot support
D. Which part of my argument it is for

32 5. Research, sources and evidence

What does 'peer reviewed' actually certify?

A. That the data were checked and the analysis recomputed
B. That the finding has been replicated independently
C. That two or three people in the field recommended it
D. That the journal is not predatory and is properly indexed

33 5. Research, sources and evidence

What is the correct use of a Wikipedia article in academic work?

A. Cite it for uncontroversial background facts only
B. Cite it once the talk page shows no active dispute
C. Paste it as the source for the model to work from
D. Read it, then use its reference list

34 5. Research, sources and evidence

A page looks professional and cites figures. What do fact-checkers do that students left to themselves do not?

A. They read the page more slowly and carefully than students
B. They check the page's design against known templates
C. They open new tabs to find out who runs it
D. They run the text through a detection tool first

35 5. Research, sources and evidence

A deep research mode returns a twenty-page report with sixty citations. What is it good for?

A. A bibliography you can submit with its reference list intact
B. A map: which areas, which names, which terms to search
C. A first draft of the literature review section
D. A verified summary of the consensus in the field

36 6. Writing, and keeping your voice

You cannot say what your argument is in one sentence out loud. Which cause of blank-page paralysis is that?

A. You do not know enough yet
B. You have not decided what you think
C. You require the first sentence to be perfect
D. You are avoiding it because it is unpleasant

37 6. Writing, and keeping your voice

Which short request removes the essay, despite producing very little text?

A. Ask me five questions that would make me more specific
B. What are the ambiguous words in this question?
C. Give me a thesis for this question
D. Which two of my topic sentences make the same point?

38 6. Writing, and keeping your voice

Which part of a paragraph do students most often omit, and where a large share of the marks sit?

A. The topic sentence stating the claim
B. The citation with a page or figure
C. The link into the following paragraph
D. Why the evidence supports the claim

39 6. Writing, and keeping your voice

You are four hundred words over the limit. What must not be cut?

A. Background the reader of your course already has
B. Sentences summarising what you are about to say
C. Quotations running longer than a single sentence
D. The objection, stated at its strongest

40 6. Writing, and keeping your voice

An editing pass has raised the register of your prose. What should you check?

A. Whether it also raised the precision
B. Whether the word count stayed within the limit
C. Whether the vocabulary matches the reading list
D. Whether the sentences are now of even length

41 6. Writing, and keeping your voice

Which source of feedback do students already hold and almost never use?

A. Comments on their last five marked pieces
B. The module handbook's stated learning outcomes
C. The writing centre, which most universities run
D. Office hours with the person marking the work

42 6. Writing, and keeping your voice

What is the tell that a presentation was built the wrong way round?

A. The slides could be read without the speaker
B. The speaker rehearsed the transitions between slides
C. The deck has more slides than minutes available
D. The figures were plotted rather than drawn by hand

43 7. Rules, honesty and the record

The university policy permits AI with disclosure and the assessment brief says nothing. What governs?

A. The university policy, as the higher-level document
B. The module handbook and, failing that, a written answer
C. Whatever the majority of the cohort is doing in practice
D. The practice on a friend's programme at another university

44 7. Rules, honesty and the record

Some policies turn an otherwise permitted use into misconduct. How?

A. By capping the number of sessions allowed per assignment
B. By requiring a particular institutional tool to be used
C. By treating proofreading as generation past a word count
D. By making non-disclosure itself the offence

45 7. Rules, honesty and the record

What does a disclosure statement not achieve?

A. It does not make a prohibited use permitted
B. It does not satisfy an examiner who wants a viva
C. It does not cover uses made by other group members
D. It does not count unless the brief provides a form

46 7. Rules, honesty and the record

An allegation arrives, based on a detector score. What is the first mistake to avoid?

A. Admitting to something you did not do
B. Asking what corroborating evidence exists
C. Contacting the students' union adviser early
D. Exporting your version history immediately

47 7. Rules, honesty and the record

What distinguishes a legitimate adjustment from an unfair advantage?

A. It is used by fewer than five per cent of a cohort
B. It removes a barrier without supplying the substance
C. It was recommended by a lecturer rather than a student
D. It applies only to coursework and never to examinations

48 7. Rules, honesty and the record

Of storage, training and human review, which is the one you can usually control?

A. Whether it is retained after you delete it
B. Whether a human can read it in an investigation
C. Whether your institution can see it on its network
D. Whether it is used to train the model

49 7. Rules, honesty and the record

What is the trade-off of using the AI tool your university provides?

A. It is slower, because the traffic is inspected
B. It cannot be used off the campus network
C. Logs exist and are available in a case
D. Its output may not be cited in submitted work

50 8. Assessment, and the years after it

Which revision activity is now very common among students with these tools and close to the worst use of the time?

A. Sitting a past paper and marking it strictly yourself
B. Having AI condense your notes and studying the summary
C. Rewriting your notes from memory and checking the gaps
D. Writing eight questions on a chapter you have just read

51 8. Assessment, and the years after it

Why is a log of everything you reached for a tool to get better than your sense of which topics are weak?

A. It is shorter and therefore quicker to revise from
B. It matches the weighting of the examination paper
C. It can be shared with a tutor for a second opinion
D. It was recorded at the moment of not knowing

52 8. Assessment, and the years after it

Why does answering the wrong command word cost the whole question rather than a few marks?

A. They are technical terms with fixed meanings
B. They determine how long the answer should be
C. They indicate which topic the question belongs to
D. They change between examination boards each year

53 8. Assessment, and the years after it

What does a weekly working document on a long project do that writing at the end cannot?

A. It shortens the final draft by reusing the same text
B. It converts reading into thinking as you go
C. It satisfies the supervisor's meeting requirement
D. It guarantees the question will not need to change

54 8. Assessment, and the years after it

What is the test for every sentence in a personal statement?

A. Could another applicant have written it?

B. Does it use the vocabulary of the course?

C. Does it demonstrate a transferable skill?

D. Would it survive a similarity check?

55 8. Assessment, and the years after it

On placement, three rulebooks apply at once. Which governs?

A. The university's, since it awards the qualification

B. The strictest of them

C. The host organisation's, but only during working hours

D. The professional body's, once you are registered with it

56 8. Assessment, and the years after it

What is the single habit the course ends on?

A. Disclose every use, however small, in writing

B. Ask what changed: the document, or you

C. Verify every specific against a source you opened

D. Keep version history on for every piece of work

Answers

Each entry gives the correct option, then what each wrong one reveals about how somebody was thinking. The second part is worth more than the first: a wrong answer here is a named misreading, not a mistake.

Lesson checks

1. Two Kinds of Help — C

A — Treats the problem as one of paraphrasing and ownership, as if learning were a plagiarism question rather than a memory one.

B — Over-reads the advice to learn from mistakes, making error the mechanism when the mechanism is effortful retrieval, right or wrong.

D — Assumes clarity while reading equals knowing, which is exactly the fluency illusion a well-written solution creates.

2. What Happens To Your Memory — B

A — Reads the result as discrediting the practice measure, when the point is that practice performance and durable learning genuinely came apart.

C — Assumes the AI group was lazier, but they performed better during practice, not worse — they were doing something else, not less.

D — Treats model capability as the variable when the manipulated variable was answer-giving versus hint-giving, which a stronger model does not change.

3. What it is actually doing — D

A — 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 — Plausible-sounding but wrong scale: a 611-word essay is roughly 800 tokens, far inside any modern context window.

C — Confuses variability with capability; at temperature 0 the model would give a stable wrong count, not a right one.

4. Why it invents things, and when — B

A — Assumes the hedging in training text tracks the model's internal probabilities; it tracks the surrounding style instead.

C — Consistency within one conversation is expected — the earlier answer is in the context and heavily conditions what follows. Only a cold rephrasing in a fresh chat tests stability.

D — Borrows a real concept, calibration, but confuses measured token probabilities with the confident tone of the output text; the chat window never shows the former.

5. The fluency trap — A

B — Treats study time as fungible; the retrieval-practice literature exists precisely because equal time produces unequal retention.

C — Substitutes a real but much smaller effect for the retrieval effect, and treats confidence as a cause of performance rather than a symptom of fluency.

D — Reasonable-sounding, but failed retrieval followed by feedback is among the strongest conditions in the literature; the attempt does the work even when it fails.

6. Context, limits, and why it forgets — C

A — Blames the file format for a retrieval problem; the same PDF page pasted alone produced the right answer.

B — Confuses training memory with supplied context; supplied text does not need to have been in training to be usable.

D — Conflates billing limits with context handling; a quota breach stops the request rather than degrading the answer invisibly.

7. Which tool for which job — D

A — The check happens inside the same predictive process that produced the answer, so a consistent-looking chain can be wrong at both ends.

B — Two agreeing predictions raise confidence slightly but share the same failure mode on unusual integrals; a symbolic engine settles it outright.

C — Steps are worth asking for and do catch many errors, but a plausible invalid step reads exactly like a valid one; visibility is not verification.

8. Getting set up for nothing — B

A — Existence questions need a database, and a small local model has weaker compressed knowledge than a hosted one, making fabrication more likely rather than less.

C — Any model without retrieval, local or hosted, has no access to recent information; the local one also has no search to fall back on.

D — Assumes parameter count does not matter for knowledge; a 3B model holds a far coarser compression of its training text than a large hosted one, and unfamiliar advanced material is exactly where that shows.

9. A working agreement with yourself — C

A — Ties a career-length decision to a document that changes annually and covers only what is examined.

B — Anchors on today's capability, which moves; the skills worth keeping are often ones the tools already do well, because you must judge that output.

D — Uses efficiency as the criterion for a decision about capability; the highest time savings come precisely where the learning loss is largest.

10. The anatomy of a study prompt — A

B — Form controls the shape of the answer; the complaint here is about where it started, which is a level problem.

C — A real weakness in the prompt and a genuine second-order fault, but naming the task would not on its own have stopped the answer beginning with the definition of a variable.

D — Explains the length but not the starting point; a 200-word answer pitched at a beginner would have the same problem in less space.

11. Give it the source — D

A — Assumes any error means the source was ignored; a summariser that drops a qualifying clause is still working from the source.

B — Treats a comprehension failure as an ingestion problem; the text was fully present either way.

C — A handout is a few thousand tokens at most, far inside any current window; the sentence was visible and was compressed away.

12. Ask for the working, then check it — C

A — Invents a mechanism; sampling applies uniformly across the output, and a stable chain can still be unfaithful.

B — Overstates the mechanism into a specific claim that is not generally true; generation is left to right, and the steps do genuinely improve accuracy.

D — Treats the model as retrieving a stored solution; it is generating one, which is why the chain adapts correctly to a novel problem's numbers.

13. Follow-ups that find the weak joint — B

A — A real effect at length, but the check fails immediately in a short chat too, so truncation is not the reason.

C — Invents a mechanism; sampling settings do not change as a conversation proceeds.

D — Describes the sycophancy problem, which is real, but it predicts change rather than the consistency that actually defeats the check.

14. Explain it at three levels — D

A — Confuses sounding advanced with understanding; unearned terminology is one of the easiest things for a marker to expose.

B — Treats level as syllabus position rather than as depth on the same idea, which is what the technique is doing.

C — Reverses the reliability gradient; the further from mainstream teaching material, the more the model is generating what an expert objection sounds like.

15. Conversations have a shape — A

B — Adds another message to the context that already contains the rejected structure, so the thing being conditioned on grows rather than shrinks.

C — Assumes the correction was lost to truncation; at thirty messages it is still present, and being present is the problem.

D — Closer to right than the others and sometimes available, but editing a thread mid-way rewrites everything downstream of the edit and is slower than simply starting clean.

16. What not to paste — C

A — Raises a real accuracy concern but answers a different question; the issue here is disclosure, not correctness.

B — Treats a settings question as the whole issue; the disclosure is improper even on an account contracted not to train.

D — Reaches for ownership when the operative rule is confidentiality and data protection; the duty is owed to the patient, not to the institution's rights in a document.

17. Diagnosing a bad answer — D

A — Assumes stability implicates the prompt; identical instructions producing identical output is expected, and says nothing about whether the instruction was at fault.

B — Reaches for a grounding failure when the diagnosis routine has already confirmed the material was supplied and specified.

C — Invents caching between separate conversations; models generate afresh each time, and a widespread misconception will appear across providers because they share sources.

18. Prompts worth keeping — B

A — True as a side effect and irrelevant to the learning mechanism; the same tokens spent on a question list would be equally cheap and far less useful.

C — Names a real hazard of long threads but the wrong one here; an interactive quiz produces a longer conversation than a single list, not a shorter one.

D — Adaptivity is a genuine bonus, but the technique works even with fixed questions; the effect comes from attempting recall before seeing the answer.

19. Getting Unstuck, Not Getting Finished — D

A — Treats struggle itself as the active ingredient, but the evidence is about struggle followed by resolution — nineteen unresolved minutes added nothing.

B — Effort-as-payment reasoning: the twenty-five minutes were real, but they do not change what reading a complete solution does to what you can retrieve later.

C — Sounds like efficient triage, but it protects the comfortable feeling of fluency and skips precisely the gap you just found.

20. Make It Quiz You — A

B — Treats number of exposures as the mechanism, which is what rereading maximises and what the testing-effect studies show is the weaker route.

C — A reasonable fear of practising errors, but errors that get corrected do not stick as errors — the attempt still does the work.

D — Attributes the gain to coverage rather than to the act of retrieval, which would predict that reading twenty answers works equally well.

21. Building a hint ladder — C

A — Treats the correction as a time saving rather than as the task; the diagnosis is the thing the exam will require you to perform unaided.

B — Raises an unrelated reliability worry; the omission's cost is pedagogical, and it applies even when the correction offered is perfectly right.

D — Borrows a genuine concern from the academic-integrity module, but nothing here overwrites the student's file; the loss is what they no longer have to think through.

22. Retrieval before review — D

A — Draws a magnitude claim from the wrong test; on the delayed test the gap is large, which is the comparison an exam resembles.

B — Invents a boundary condition; the testing effect is found across factual, conceptual and procedural material alike.

C — Assumes a well-run technique wins everywhere; the immediate-test advantage for re-reading is a normal, expected finding in this literature, not a sign of error.

23. Spacing and interleaving — A

B — Attributes the effect to extra effort per item rather than to what that effort is spent on; the studies hold time roughly constant.

C — Plausible and partly true, but interleaving still wins in experiments where both conditions are equally spaced, so recognition is doing independent work.

D — Substitutes a general engagement story for a specific mechanism, and does not explain why the interleaved group performs worse during the session.

24. Explaining it back — B

A — Reframes a knowledge measurement as an emotional one; the exercise is diagnostic precisely because the gaps are informative.

C — Describes a real contextual effect but a secondary one; the damage is done at the moment the student reads the answer they were about to generate.

D — A general reliability caution rather than the reason for this clause; the instruction would still be needed if every filled gap were perfectly correct.

25. Worked examples, and when to stop using them — D

A — Substitutes a motivational story for a capacity one; the effect appears even when students are fully engaged and trying.

B — Interference is real when methods genuinely differ, but the reversal effect appears even when the example uses exactly the student's own method.

C — Correctly names retrieval but overgeneralises it into the sole mechanism, which would wrongly predict that worked examples never help beginners either.

26. A study session that works — C

A — A real secondary benefit, and the module names it as such, but not what distinguishes these blocks from the log line that follows them.

B — Twenty minutes apart is far too short to be spacing in any meaningful sense; within-session repetition is close to massed practice.

D — Frames the value as abstinence rather than measurement; the blank page would be just as necessary for a student with no access to any tool.

27. Study groups, with and without the machine — A

B — Treats a procedural detail as the mechanism; ratings help the discussion start in the right place but are not what produces the gain.

C — Overreaches from one finding to a much broader claim; the discussion works on reasoning the group already partly possesses.

D — Inverts the design principle: questions targeting a strong shared misconception are supposed to catch nearly everyone, which is what makes the discussion productive.

28. Proving it to yourself — D

A — Reads the prediction as a goal rather than an estimate; the point is measurement of judgement, not motivation.

B — Treats a diagnostic instrument as emotional preparation, and would work equally well with a deliberately pessimistic guess, which would destroy the measurement.

C — Names a genuine anchoring risk in step four, but the prediction's value persists even when a marker, not the student, does the marking.

29. Maths, and the arithmetic problem — B

A — Re-runs the same estimation process that produced the slip; the model has no column-arithmetic procedure to re-execute more carefully.

C — Units are an excellent check and would catch a joules-versus-watts error, but a factor of ten leaves the dimensions entirely intact.

D — Cold re-asking tests stability of knowledge, but arithmetic errors are not stable in a useful way — you may simply get a second, differently wrong number.

30. Proofs, and the step that looks fine — C

A — Invents a positional claim; unjustified steps appear wherever the difficulty of the argument sits, which is usually in the middle.

B — Confuses a reading order with a logical transformation; nothing about reading upward changes the statement being proved.

D — Treats the conclusion as an anchor, but a correct conclusion is exactly what an invalid proof usually reaches; its correctness proves nothing about the chain.

31. Code that runs and is wrong — D

A — Often true in practice, but a thorough generated suite would have the same structural flaw; coverage is not the issue.

B — States a real test-first principle too absolutely; a human writing tests afterwards against the original specification still provides an independent check.

C — Answers a marking-policy question rather than a correctness one; the structural problem exists whether or not tests are assessed.

32. The lab sciences, and the dimension check — A

B — Confuses precision with dimension; rounding never changes what physical quantity a result represents.

C — Treats a dimensional mismatch as a units conversion; newtons and joules differ by a length, so no conversion factor exists between them.

D — Invents a convention; force times distance is energy precisely because the dimensions differ by a length, which is what the missing factor reveals.

33. History, and the quotation that was never said — B

A — Citing the repetition rather than the source shifts the blame without checking the claim; the attribution is what is in question, not who repeated it.

C — Possible in principle and a reason to look further, but treating an absence of any earlier trace as evidence for an unseen manuscript inverts the burden of proof.

D — Overstates a real limitation; nineteenth-century printed material is among the best-covered periods, and a genuinely circulating quotation would leave traces.

34. Literature, and the text it half remembers — D

A — Overstates the accuracy problem; on well-known works the thematic summary is generally serviceable, which is precisely what makes it dangerous.

B — Shifts to a detection claim that the course elsewhere shows is unreliable; the damage here is to the quality of the reading, not to how it looks.

C — A real hazard covered elsewhere in the lesson, but it applies to quotations at any stage, not to the ordering problem being asked about.

35. Languages, where it is genuinely excellent — C

A — Denies a benefit the lesson explicitly credits; controlled-level conversation does build production, provided errors are surfaced.

B — Misplaces the fault in level setting; the described problem persists at a perfectly matched level.

D — Swaps a diagnosis for a preference about time allocation, and understates the role of extended input in acquisition.

36. Statistics for coursework — B

A — Blames capability for an underspecification failure; given the full description the recommendation would likely have been sound.

C — States a contested position as settled; analysing Likert items as continuous is defensible in some circumstances and is a choice to justify, not an automatic error.

D — Substitutes a blanket power objection for the specific mismatch; n of 9 and 34 supports several appropriate analyses.

37. The essay subjects, and what a marker is reading for — A

B — Caricatures the criterion; what is rewarded is a defensible position, not assertiveness, and an indefensible strong claim scores worse still.

C — Attributes the penalty to detection; the formulation scored badly for decades before any of these tools existed.

D — Generalises one common essay form into a universal requirement; plenty of questions ask about mechanism or meaning rather than relative weight.

38. Research, And Sources That Exist — D

A — It will confirm, at length and convincingly — this is precisely the step that got a New York lawyer sanctioned in 2023.

B — The format looks institutional and authoritative, but a model can emit a DOI-shaped string that resolves to nothing at all.

C — Real researchers are routinely attached to papers they never wrote, so verifying the people leaves the paper unverified.

39. How a citation gets invented — D

A — Invents a failure mode; DOIs are persistent identifiers and are not reassigned, which is the point of the system.

B — Describes an access obstacle rather than the verification gap; the same problem would remain if every paper were open access.

C — Names a formatting issue that resolution actually does expose, since the landing page shows the correct metadata to compare against.

40. Finding the paper, for nothing — C

A — Invents a correlation; whether a cited work is open depends on its publisher, not on having been cited.

B — Reverses the coverage relationship; major databases index far more than any single paper cites.

D — Names a real obstacle the lesson raises but attaches it to the wrong solution; reading reference lists is in fact how you acquire that vocabulary.

41. Reading a paper without reading all of it — D

A — Attributes objectivity to figures; how data are plotted, scaled and binned involves plenty of interpretation.

B — Half right about sequencing but wrong about the reason for the figures' position, and it wrongly implies methods are always read last.

C — Offers a publishing-process claim that is often true and irrelevant here; the introduction's ordering is about triage cost, not sincerity.

42. What counts as evidence — B

A — Correctly demands uncertainty but overreaches: intervals would still not license comparing across trials with different populations and measures.

C — Misapplies the randomisation concept; each trial may be perfectly randomised internally, and the flaw is in comparing across them.

D — Raises a legitimate outcome-selection question that would still leave the cross-study comparison invalid even with an ideal outcome.

43. When it searches the web — A

B — Plausible on average and beside the point; the failure described is correlation between sources, not the error rate of each.

C — Makes an unsupported ranking claim and would, if true, reduce rather than create the problem.

D — Describes a real inconvenience, but generated pages frequently do carry citations, which is part of why they are convincing.

44. Notes that survive the term — C

A — Treats brevity as the benefit; a short copied note is just as unencoded and just as risky.

B — A genuine third benefit the course mentions elsewhere, but not one of the two the lesson pairs here.

D — Combines two plausible-sounding effects, but structure-finding is a consequence of paraphrasing rather than the problem it solves.

45. Building a bibliography you can defend — D

A — States a rule that varies by institution and misses the underlying reason, which would hold even where citation is permitted.

B — True and insufficient: a reliably correct but unretrievable source would still fail the test that makes citation meaningful.

C — Mistakes a formatting question for a substantive one; most style guides do now provide a form, and unpublished material is citable in other cases.

46. Misinformation, and what you pass on — B

A — Asserts a ranking behaviour that is not established, and locates the problem in retrieval rather than in the sources themselves.

C — Describes a change in reading behaviour rather than in what the sources are, and the check would still work if the underlying pages were independent.

D — Names a real capacity problem, but slow correction is a different failure from a lack of independence between the agreeing sources.

47. Writing With It Without Losing Your Voice — B

A — A real risk, but a separate one — a perfectly sourced AI draft would still have cost him the thing that actually went missing.

C — Reduces voice to surface style, which is exactly the layer rewriting does repair, while the structural layer underneath is untouched.

D — Makes the whole question about rule compliance, which is the frame that hides what happened to his own thinking.

48. Where the thinking happens — C

A — Treats style as the residue; if the argument and structure are the student's, style is recoverable through editing.

B — Confuses evidence with substance, and a version history would in fact show the outline preceding the draft.

D — Names something real that the lesson concedes drafting contains, but treats a marginal loss as the decisive one when the major stages were retained.

49. The blank page, without outsourcing it — D

A — Applies the right fix to a different cause; a placeholder helps when you know the argument and cannot phrase the opening.

B — Sometimes true and worth checking, but a student who has read enough can usually state a provisional argument even about an ambiguous question.

C — Free-writing surfaces a position you hold implicitly; it produces nothing when the underlying reading has not been done.

50. Structure that carries an argument — A

B — Mistakes navigability for argument; signposting should mark dependencies, not compensate for their absence.

C — Reads self-containment as quality; a paragraph can be internally sound and still contribute nothing to a chain.

D — Reaches the right diagnosis of description through a false claim: analysis is precisely what creates dependency between paragraphs.

51. Editing your own prose — B

A — States a style preference as a rule; passive constructions are standard and appropriate in many methods sections.

C — Uses length as the test; a longer sentence that added precision would be an improvement.

D — Names a genuine trend in style guidance but treats the pronoun as the loss, when the sentence would still be worse with "we" restored and "40" gone.

52. Your voice, measurably — D

A — Assumes a rubric line that may not exist; the disadvantage operates through the reader's judgement of understanding, not a stated deduction.

B — Confuses technical terms, which do need defining, with general vocabulary inflation, which is a different fault entirely.

C — Borrows a detection claim the course rejects elsewhere, and the problem long predates any of these tools.

53. Feedback that is actually useful — C

A — Raises a real caution about accepting edits, which applies equally whenever sentence feedback is sought.

B — Sensible time management that does not explain the ordering; the reason is that structural changes destroy sentence work, not that they are slow.

D — Plausible claim about marking that is beside the point; the ordering would hold even if both carried identical weight.

54. Presentations, posters and visuals — A

B — Raises an unsettled legal question that applies to both images equally and is not what distinguishes them.

C — Grounds the distinction in marking rather than in truth; a wrong diagram would still be wrong in unassessed work.

D — Correct about data figures specifically and an over-extension here; an accurate hand-drawn mechanism diagram needs no dataset, so provenance is not the operative fault.

55. Group work, and shared rules — D

A — Frames it as efficiency; the rule would still be necessary for a list one person could check alone.

B — Invents a common requirement; declarations vary widely and many groups submit none, while the underlying hazard is present either way.

C — Treats the rule as apportioning blame after the fact; joint submissions implicate everyone regardless of who wrote what.

56. The Rules, And Why Detectors Do Not Work — C

A — Reads the score as a calibrated probability of guilt, which is how such numbers get used in hearings and is not what they measure.

B — Imagines a plagiarism-style corpus match; detectors compare nothing, they score statistical predictability of your own words.

D — Sounds sophisticated and even generous, but the tool has no way to distinguish stylistic influence from anything else it is measuring.

57. Reading your actual policy — B

A — Describes a different problem, and disclosure requirements do not in themselves reverse any burden.

C — Confuses disclosure with prior permission; a disclosure regime allows the use and asks you to state it.

D — Voices a common fear rather than the structural trap; the hazard is the offence being committed by silence, not by the marker's reaction.

58. Writing a disclosure that helps you — C

A — Asserts a specific penalty structure that varies by institution, rather than the general point about what the document becomes.

B — Relies on detectors as a reliable comparator, which the module explicitly rejects.

D — Speculates about the marker's inference; the difficulty is created by the document itself being false, regardless of how it is interpreted.

59. How detectors work, and why they cannot — D

A — Raises a real external-validity concern that would make things worse, but the arithmetic already fails at the stated accuracy.

B — Uncertainty in the prevalence widens the range but does not remove the effect, which holds across every plausible value.

C — Overstates into a claim that no probabilistic evidence is ever informative about a case; the point is the size of the resulting probability, not its inapplicability.

60. The process record — A

B — Invents an evidential rule; dated photographs of drafts are ordinary and useful evidence.

C — Conflates two distinct signals; the problem is the absence of development, not a rhythm anyone is measuring in a normal case.

D — Late work is common and unremarkable on its own; the weakness is what the history omits, not when it occurred.

61. If you are accused — C

A — Overstates: a conversation can be prepared for to a degree, and the module recommends re-reading your work before the meeting.

B — Invents a general evidential hierarchy; procedures vary and most consider all evidence together.

D — Rests on a false inference about motive; capable students use these tools too, and the question is authorship of this piece.

62. Where the line really is — B

A — Assumes procedural practice rather than reasoning about the principle; panels apply written policy, which is often more specific.

C — Draws the wrong conclusion: the question guides judgement where policy is silent, and never replaces reading the policy.

D — Not reliably true in either direction; the question sometimes permits things a specific policy forbids, which is why the policy still governs.

63. Accessibility, fairness and who these rules fall on — D

A — Registration matters evidentially and does not determine the substance; a registered student can still cross the line, and an unregistered one can use a tool legitimately.

B — Provenance of the software does not settle how it is used; a permitted tool can be used to generate a submission.

C — Close to right and phrased as though criteria settle it; the substance test applies even where criteria are vague or unwritten.

64. What happens to what you type — C

A — Conflates training with retention; conversations are typically kept for a period for abuse monitoring regardless of the toggle.

B — Misattributes the purpose of review, which is largely safety and abuse monitoring rather than training curation.

D — Assumes a capability that does not exist; weights already trained cannot be selectively unlearned on request, which is why the timing of the setting matters.

65. Revision That Survives The Exam Hall — B

A — Coverage feels like progress, but this recreates exactly the unguarded-tutor condition that raised practice scores and lowered exam scores.

C — Feels efficient and produces a strong sense of understanding, but the compression was the thinking, and it was done by something else.

D — Correctly identifies breadth as valuable, but reading answers is not retrieval — the questions only pay off if you attempt them cold.

66. The Honest Case For Still Learning It — D

A — The historical parallel is genuinely strong and partly correct, which is why it needs the output-versus-judgment distinction to be answered rather than dismissed.

B — True today, but it makes the case depend on current error rates — the argument would collapse the moment a model became reliable, and it should not.

C — Concedes the practical ground and retreats to the aesthetic case, which is exactly the move that loses the argument with a sceptical student.

67. Past papers and mark schemes — A

B — Raises accuracy, which is a real risk and not the main cost; a perfectly correct model answer would produce the same problem.

C — Names a real mismatch that pasting the actual mark scheme would fix, leaving the deeper problem untouched.

D — Imagines a plagiarism risk in a closed exam written from memory, where the realistic cost is simply not having practised.

68. Practising under the conditions of the test — D

A — Names a genuine cost the lesson raises, but a paper of the same questions would still be easier with the tool available.

B — Describes a likely outcome for one group of candidates rather than why the assessment itself is not easier.

C — Invents a technical constraint; institutions running such exams provide reliable access.

69. Vivas, and being asked about your own work — B

A — Nearly right about location but wrong about why it works; someone could ask a model for plausible alternatives, and the giveaway is being unable to say why each was unsuitable for this case.

C — Reframes the question as a coverage check; the examiner is probing judgement about this specific piece of work.

D — Turns a diagnostic signal into a rubric line; naming alternatives is easy, and explaining why you set each aside is what carries the information.

70. Long projects and dissertations — C

A — Invents a compliance check; scope approval is not usually re-examined against a project's history.

B — Attributes it to a marking criterion; the value is that you can explain the reasoning, whether or not anyone assesses the change itself.

D — Treats change itself as evidence of quality; a question that drifted without reasons is a symptom of the problem, not of engagement.

71. Applications, statements and interviews — D

A — Relies on detection, which is unreliable and is not why the document fails to work.

B — Assumes a reader is comparing registers forensically; the statement would fail on a first reading purely for being interchangeable.

C — Not strictly true — a model will happily include experiences you supply — and it locates the fault in capability rather than in what generation optimises for.

72. Placements, and rules that are not your university's — A

B — Overcorrects into a rule that would forbid ordinary case-based learning; the general clinical question can be discussed without the particulars.

C — Invents a procedural timing rule and treats a disclosure question as the operative one.

D — Names a retention risk that is real but secondary; the disclosure to a third party is complete at the moment of pasting, whatever happens to the logs afterwards.

73. What to learn anyway — B

A — Bets on scarcity or failure of access, which is a weak and probably wrong prediction.

C — True in the short term and contingent on hiring practice, which is exactly the sort of thing that changes.

D — Leans on broad transfer, which the evidence supports much less than people assume, and which the course elsewhere treats sceptically.

Module test — 1. What the machine is doing when it helps you

1. A

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.

2. C

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.

3. B

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.

4. D

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.

5. A

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.

6. C

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.

Module test — 2. Asking well

1. A

The map is one to one and diagnosing takes five seconds. Too basic means you left out level. An answer that wandered means you named a topic instead of a task. Too long means no constraints. The wrong shape means no form. The highest-value addition is a statement of your current, possibly mistaken, understanding, because it converts a general explanation into a diagnosis aimed at your actual error.

2. B

Grounding removes the invention failure and leaves two others. It can misstate what the passage says, particularly for negations, conditions and exceptions — unless the sample is randomised is exactly the clause that goes missing. And if your notes contain an error, or you pasted last year's version, it will faithfully explain the error. So the question stops being does this exist, which needs a database, and becomes does the passage actually say that, which you settle in ten seconds by looking.

3. D

A model that has just written an answer will tend to be consistent with it, because the conversation is a block of text it re-reads in full before every reply. Consistency inside a conversation measures almost nothing. Two cold rephrasings that agree are weak evidence that the knowledge is well represented; two that disagree are strong evidence that you need a real source, and it costs one message.

4. C

Drift, anchoring, compounding agreeableness and silent truncation all follow from one fact: the model conditions on the entire thread, including the parts you abandoned. Correction works less well than starting again, because the wrong version is still sitting there being conditioned on. Ask for the conclusions and constraints in under a hundred and fifty words, check that summary is right, and paste it into a new window — fifteen seconds, and you keep the signal and drop the noise.

5. A

A lecture slide, a textbook page or your own draft harms nobody if it ends up in a public archive with your name on it, so paste generously — supplied text beats remembered text. A classmate's draft, a placement case or participant data belongs to somebody who did not choose it, and a setting is a promise about handling rather than a guarantee about outcomes. Removing names is not de-identification when a rare condition, a small town and a date identify somebody in combination.

6. A

Generating intermediate steps genuinely does improve performance, and a chain gives you something to inspect where a bare conclusion gave you nothing. What it does not give is faithfulness. Research on this repeatedly finds stated reasoning that does not match what drove the answer, including models given a hint that change their answer to match it while producing a chain which never mentions the hint. Treat visible working as a document to audit, not as verification.

Module test — 3. Study that raises the difficulty

1. C

The act of pulling something out of your head changes how available it is next time, so a quiz is not a check on studying, it is studying. The awkward part is the confidence: the re-readers had predicted they would do better, and they were wrong. Every incentive in the moment points the wrong way, which is exactly why an assistant that explains fluently is so easy to mistake for study.

2. A

Which line is wrong points you at the error; what it should be removes the work, and the gap between locating an error and being told the fix is where the learning happens. This is the clause people leave out of the hint-ladder prompt and it is the most valuable one. If you still cannot see it after a genuine attempt, escalate to the kind of error rather than its content: algebraic, conceptual, or an unjustified assumption.

3. D

Choosing the method is a separate skill from executing it, and blocked practice never exercises it, because the heading already made the choice. An exam paper does not tell you. That is why blocked practice wins the session by a wide margin and loses the later test by a similar one, and why the clause that matters when asking a model to quiz you is without telling me which topic it is from.

4. B

A beginner attempting an unfamiliar problem spends a small working memory entirely on searching, so an example removes the search and attention goes to the pattern. Somebody who already has the method must process an explanation of something they can do, which costs capacity and returns nothing. The answer is neither examples nor problems but a fade between them, removing steps from the end — and a decision in advance about how many examples you will read, because feeling ready is not a reliable signal here.

5. C

The average gap between predicted and actual is the most useful single statistic you can hold about yourself. If you run fifteen per cent hot, you now know when to stop revising a topic and when you were wrong to, which is worth more than any individual mark. Students leaning hardest on an assistant tend to over-predict most, and measuring your own gap is how you find out whether that describes you rather than assuming either way.

6. A

Students answer alone, discuss with somebody who chose differently, then answer again, and the proportion correct rises substantially — including in groups where nobody initially had it right, which shows the gain comes from articulating and challenging reasoning rather than from copying the strongest student. Consulting the assistant first supplies the answer and removes the exercise. The reasoning is the thing being trained, and it is the first casualty of a shared tool.

Module test — 4. Subject by subject

1. D

Depending on the tokenizer, 4763 might arrive as 47 and 63, so there is never a column of digits to carry between. The model produces the most plausible continuation, which for small familiar numbers is almost always right and degrades steadily with length. The failures look exactly like the successes — a confident wrong number in the middle of correct working — so take the method from the model and the number from SymPy, Maxima or a code interpreter that actually computes.

2. B

Models produce proof-shaped text extremely well, so the failure is not gibberish. One unjustified step makes the whole thing worth nothing, however correct the conclusion happens to be, and markers find these immediately because an unjustified step is exactly what they are trained to look for. Read upward from the conclusion asking what licenses each line: forward reading rides the momentum of a plausible narrative, and if you cannot name the rule, the step is not proved.

3. C

Tests produced by whatever produced the code pass by construction, because they encode the same understanding of the problem — so they check the code's assumptions rather than the specification's. Write the tests first, yourself, from the spec, then let the model help with the implementation; that order preserves the check. It is the same reason generated code that runs is not generated code that is correct, and why you test the boundary: empty input, one element, a negative, a duplicate.

4. A

Units carried through as algebra must cancel to the unit of the answer, so a mismatch localises the fault upstream and you can stop reading there. The check takes about fifteen seconds and catches a startling proportion of errors, because a wrong formula usually has wrong dimensions — kg m s^{-2} where you wanted kg m^2 s^{-2} is a v that should have been v squared. It also lets you reconstruct a formula you have forgotten, up to a constant.

5. D

Famous figures attract invented quotations, and the misattributions are themselves in the training data, so the model reproduces them confidently. A full-text search of published works restricted by date shows when a phrase first enters print, which exposes most inventions in about two minutes, for nothing. It is the bibliographic half of the subject's own discipline: every quotation gets traced to a primary source or a scholarly edition, or it does not get used.

6. B

Design, measurement level, sample sizes and the actual question determine the test, and measurement level is the one most often left out and the one that changes the answer completely. A five-point questionnaire item is ordinal, and a t-test on it is a choice that needs defending. Supply all four and the recommendation is usually sound; leave one out and the model fills the gap with a default, which is how a t-test ends up on ordinal data with seven people in a group.

Module test — 5. Research, sources and evidence

1. C

A real DOI lands on the article and an invented one errors, which disposes of most fabricated references in ten seconds. That is the cheap half of two required checks. The expensive half is opening the paper and finding the sentence, because the dangerous failures are hybrids: a real paper attached to a claim it does not make, or one whose finding is reversed in your sentence. Every other element of those checks out and the citation is still false.

2. A

Retrieval genuinely reduces pure invention and replaces it with something subtler and easier to trust by mistake. Independent evaluations keep finding a substantial minority of citations that do not support the sentence they are attached to. Retrieval also optimises for what ranks rather than what is reliable, summaries drop conditions and hedges, and a growing share of the web is itself generated, so agreeing pages are weaker evidence than they were. The check is unchanged: open the link and find the sentence.

3. B

Abstracts are written to be read and cited, they drop the qualifications, and the tendency for abstract conclusions to be stronger than the results support has been measured in several fields. If a claim carries weight in your essay, one minute in the results settles it. The reading order that works puts figures and captions second and methods last, and stopping after the abstract is a successful outcome rather than a failure.

4. D

Randomisation means the groups differ on average only in the intervention, which is what causal language rests on. A cohort study establishes temporal order, a real gain, and stays open to confounding you never measured. A cross-sectional survey cannot tell you which way the association runs. And comparing across two separate trials — sixty per cent response in one, forty-five in another — is invalid however natural it reads, because the populations, measures and conditions differed.

5. A

A note is good if, three months later, it reminds you of something you understood. Deciding which three to five ideas would let you reconstruct the rest is the encoding, and a summary you did not write is a summary somebody else's process performed — what you retain from reading it is a feeling of familiarity with no structure of your own to hang anything on. Writing in your own words also removes the accidental-plagiarism risk, because three weeks later you cannot tell which sentences were yours.

6. C

A citation exists so that a reader can go and check. A conversation nobody else can open is not retrievable, so it supports nothing, which is why most style guides treat it as software or as personal communication rather than as a source. You declare the use as part of your method, and the fact underneath still needs a real source that you opened and can give a page number for.

Module test — 6. Writing, and keeping your voice

1. B

If the model drafts and you edit, the argument is already made — which claims appear, in what order, what is left out — and rewriting every sentence leaves an essay you did not write, because the essay is the argument rather than the sentences. If you draft and the model attacks, you keep the argument and gain a reader who responds immediately at two in the morning. Same tool, same polish, completely different piece of work.

2. D

Ideas are half-formed until you try to put them in order, and the attempt is what finishes them: you discover in the third paragraph that two points you thought were separate are the same point. Help at editing costs almost nothing, because the argument is fixed by then. Help at the claim and the ordering removes the thinking itself, and no amount of later rewriting makes the piece yours, because it is now a defence of somebody else's claim.

3. A

A real argument has dependencies: paragraph three establishes something paragraph four uses, so reordering breaks it and paragraph four now refers to something that has not happened. A list survives reordering because each part stands alone. Reorderability is the clearest structural signature of generated prose, since it is assembled as a set of relevant sections rather than as a chain, and it is also the difference between the middle band and the top one in most rubrics.

4. C

If you cannot say why the new version is better, do not take it. That one rule keeps the prose yours and turns each edit into a small lesson in writing. Ask for diagnosis rather than replacement, because a rewritten paragraph teaches nothing — you cannot see the reasoning that produced it. And reject the standard drift by default: nominalisation, hedge inflation, elegant variation, flattened sentence length, and specific examples replaced by general statements.

5. D

It converts a vague quality question into a rubric-matching task on supplied text, which these systems do well. Ask which band each criterion currently sits in, with the sentence that decides it quoted, then the three changes that move the most marks, and forbid rewriting and encouragement. It also produces no text for your submission, so it is permitted under nearly every policy. What to ignore is a predicted grade: the model knows nothing about your institution, cohort or marker.

6. B

A generated diagram is a picture of what a diagram looks like, carrying plausible wrong labels, arrows in the wrong direction and structures that do not exist. Draw it in Inkscape or draw.io, or plot it from data. Decorative images are usually fine and should be declared where the institution asks. Anything representing your own data must come from your data: generating or beautifying a results figure is research misconduct rather than a style choice, and that line is not blurry.

Module test — 7. Rules, honesty and the record

1. C

The two usual signals are perplexity — roughly how predictable each word is given the ones before it — and burstiness, the variation in sentence length and complexity across a document. Generated text tends to be low on both. So does careful, plain, formally structured prose by somebody writing in a second language. The instrument is measuring language proficiency and reporting it as authorship, which is not a flaw to be patched in the next version but what the measurement is.

2. A

Fifty essays are generated and forty-seven or forty-eight of those are flagged. Nine hundred and fifty are human, and five per cent of them — about forty-seven — are flagged too. So about half of everyone flagged has done nothing, and that is with generous accuracy assumptions and a fairly high prevalence; lower either and it gets worse fast. This is the arithmetic that governs medical screening, and a screening result is not a verdict.

3. B

Unread sources is the actual offence in most cases, so the verification claim closes the question the rest of the statement only opens. A disclosure is a factual account rather than a confession: which tool, at which stage, what happened to the output, and what is unambiguously yours. The test of a good one is whether a marker could tell exactly what you did without needing to ask a follow-up question.

4. D

Version history records the shape of a document arriving, not the work being done. Two thousand words appearing in one action at 23:40 tells a story that is hard to contradict, even when every word was written by hand in a notebook. The repair is cheap and has to happen in advance: photograph the pages, dated. A real writing history is messy, and that messiness is the signal — which is also why simulating one is a separate and more serious offence.

5. C

A translation exercise measures your ability to translate, so a translator defeats it entirely. An essay measures your ability to construct an argument, so corrected grammar does not touch it and a constructed argument destroys it. A programming assignment measures whether you can write the program, so an explained error message is fine and a written function is not. The question generalises to tools that do not exist yet, which no list of rules does.

6. B

Low perplexity and low burstiness come from common words, standard constructions and even sentence length, which is what a careful learner of a language and a student taught a fixed essay shape both produce. Published evaluations have found a majority of non-native speakers' essays misclassified while nearly all native-speaker essays were classified correctly. The same logic exposes autistic students writing in a consistent register, and the only available protection is a record kept in advance.

Module test — 8. Assessment, and the years after it

1. A

Past papers tell you what is actually asked, in what form, with what weighting and where the marks sit, so they are study material rather than the assessment of it. Doing one badly in week three produces information while you can still act on it. Do the question timed and closed, mark it yourself against the scheme first, and only then ask which marks were lost and where — never for a model answer, which is the fluency trap in its purest form.

2. C

Material learned in one context is retrieved more easily in that context and less easily elsewhere. If every hour of revision happens in a comfortable chair, typed, untimed, warm because you have just read about the topic, and with a tab open that answers, you have practised a different task from the one being assessed. Closing each gap is free: a timer, blank paper, a silent room, a phone in another room, and a cold start.

3. D

The document records the destination, not the route, so what a line does can be read off by anybody competent. What you tried and abandoned exists only in the history of the work, which makes it the strongest signal of ownership and the hardest thing to fake. Examiners are finding the edge of your knowledge because that is their job, and a calm I do not know, but I would look at X scores better than a confident invention.

4. B

The gap is invisible until the one closed component, and by then it is far too late to close. The protection is the unaided check, run regularly all year: close everything, set a timer for the length the real task would take, do it, then mark it honestly and write down the specific failure rather than a grade. It costs about an hour a month and it is the difference between finding out in week five and finding out in the hall.

5. C

Output can go, and pretending otherwise makes the whole case look dishonest: long division of six-digit numbers by hand, log tables, memorising a bibliography format. Judgement cannot, because there is nothing above judgement to check it, and supervision is the work that is left wherever these tools are used well. The usable form of the question is what will I no longer be able to do if the machine does this — sometimes the honest answer is nothing I need.

6. A

When everybody has the tool, the questions change: judgement, application to a novel case, critique of a supplied answer, defence of a choice. These papers are not easier. Under a clock the tool is frequently slower than knowing the answer, and knowing which is which is part of what is being tested, so practise with it under time pressure and practise catching its errors. And find out which kind of paper you are sitting rather than assuming.

AI for Students — final examination

1. A ***1. What the machine is doing when it helps you***

Every assignment has one thing it exists to build, and everything else is scaffolding that is fair game. In a problem set it is the solving, in an essay it is the argument, in a lab report it is the interpretation, in a translation exercise it is choosing between two defensible renderings. If the hour saved on a method section you have written eleven times goes into the interpretation, the tool has done its job.

2. B ***1. What the machine is doing when it helps you***

Clarity produces the sensation of understanding before any of the work that earns it, and a model raises fluency more than a textbook can, because it will rewrite the explanation until it feels effortless. The gap between what you can read along with and what you can produce cold is the whole subject, and it is precisely the gap between the tutorial and the exam. The trap bites hardest on material that is new, mechanical and multi-step.

3. D ***1. What the machine is doing when it helps you***

A token is roughly four characters of English, and the vocabulary was built mostly from English text, so the same sentence in another script can cost two or three times as many tokens. Providers charge and limit by tokens because tokens are what the computation costs, so an allowance empties faster. Nobody chose this as policy; it falls out of how the vocabulary was built, and it is one of the quieter inequalities in the technology.

4. A ***1. What the machine is doing when it helps you***

A small local model hallucinates more, follows multi-step instructions worse and gives up on long documents. For drilling, rephrasing and generating practice questions from text you supply it is entirely adequate, and it needs no connection, has no quota and keeps everything on your machine. Knowing which of those two categories a task falls into is the skill. Ollama and llama.cpp are both free, and LM Studio gives a graphical interface if terminals are unwelcome.

5. C ***1. What the machine is doing when it helps you***

Supervision is most of what is left of many jobs, and you cannot supervise a skill you never acquired. The list should be five or six capabilities rather than subjects, because a long list is one you will abandon, and everything not on it is fair game. The point of writing it down calmly is that the hard calls arrive at eleven at night with a deadline tomorrow, and nobody decides well there.

6. B ***1. What the machine is doing when it helps you***

Most chat interfaces sample somewhere in the middle of the temperature range, so the same question can produce two different answers and occasionally one right and one wrong. Two independent agreeing answers are weak evidence; two that disagree are strong evidence that you must go to a source. It only works cold, with the question rephrased in a new chat, because a model that has just written an answer conditions heavily on it.

7. C ***1. What the machine is doing when it helps you***

The prose around the specifics is usually fine. The specifics are where the weakly-represented material lives, and they are also the parts a marker checks. Marking them takes about ten seconds and catches nearly everything that would otherwise cost marks. Asking which part it is least confident about is worth doing as a list of things to verify, because it points at what is unusual or contested, but it is not a confidence report and must not be treated as one.

8. D ***2. Asking well***

Level three names what the simplification hid, where the standard treatment breaks and what a specialist would object to, and it is the pass almost nobody asks for. The distinction between a true statement and a teaching approximation is frequently exactly what separates a good answer from an excellent one. It is reliable on mainstream material and not on the frontier of a narrow field, where the right use is to find out what to ask a human.

9. B ***2. Asking well***

Everything in the current context conditions the next token strongly, so a model that has just written an answer tends to defend it. Stripped of the context in which it produced the answer, it critiques much more freely. It will sometimes invent objections, and those are cheap to evaluate and it catches real errors often enough to be worth one message. What it will never do is report its confidence, which is why are you sure has no place in the list.

10. D ***2. Asking well***

Counting, exact arithmetic, alphabetising, checking whether something exists and knowing what happened last week are category errors, and no phrasing rescues a category error. The other three each have their own fix: the four elements, a new chat carrying a short summary, and supplying the material. The most experienced users are conspicuously quick to close the window and open something else, and that speed is most of what makes them look skilled.

11. A ***2. Asking well***

Without one at a time and wait, you get a list of ten questions and ten answers, which is reading. The clauses doing the real work in every saved prompt are the restrictions: one at a time, do not rewrite, do not help me, never give the full solution unless I type SOLVE. A saved prompt then fails in one predictable way — it keeps working after it should have stopped — so read it before pasting and change the clause that no longer fits.

12. A ***2. Asking well***

Patient details do not leave the systems they are meant to live in, and pasting them into a chat window is a disclosure to a third party regardless of intent, regardless of whether names were removed, and regardless of what the university's AI policy says. De-identification fails when a rare condition, a small town and a date identify somebody in combination. Discussing the mechanism or the guideline is legitimate and is what a textbook would have given you anyway.

13. B ***2. Asking well***

Asking which assumption, if wrong, would change the answer most points straight at the step worth reading properly. In coursework the assumption list is frequently worth more than the answer, because it is exactly what a marker is looking for and it is the part students omit. The same move works outside the sciences: set the argument out as numbered premises and a conclusion, then ask what each premise can actually support.

14. D ***2. Asking well***

Standing instructions are also standing context, so a note about your level quietly conditions every future answer, including on topics where you now need the full treatment. Revisit them. The second warning is that whatever you write there is stored on the provider's servers alongside any memory feature, so keep it professional and impersonal. Boilerplate belongs there; the six prompts you actually reuse belong in a plain file you can copy from.

15. C ***3. Study that raises the difficulty***

Not understanding what is being asked, and missing a fact or a formula, are lookup problems, and nobody builds character by not knowing what a word means. Knowing that the integral of one over x is the natural log of the modulus of x is lookup, not insight. The learning lives in the other two: knowing the method while the execution goes wrong, and having an answer that is wrong without seeing why, which is where the most learning is and the one people give away fastest.

16. A ***3. Study that raises the difficulty***

The productive-failure literature is about attempt, failure, then teaching. Struggle alone, with no resolution, teaches nothing. A workable rule is that six or seven minutes with no forward progress has finished producing value — not the six minutes spent slowly assembling something, but the six spent rereading the same line. Take the smallest hint at that point, and then redo the whole thing from a blank page later the same day.

17. D ***3. Study that raises the difficulty***

Generating something yourself makes it stick harder than reading the same thing, which is the generation effect. And writing a good question requires knowing which of the chapter's ideas were load-bearing, which is a stiffer test than answering one. The model's job afterwards is to tell you which important ideas you failed to ask about — finding your blind spots rather than doing the thinking for you.

18. B ***3. Study that raises the difficulty***

You are asking a machine to quiz you on material you have not learned, which is exactly the situation where you cannot catch its mistakes. It gets worse the further you are from big, well-covered subjects: introductory biology in English is mostly reliable, a national curriculum's treatment of a local legal code in a language with less material online much less so. When a correction surprises you, check the textbook before you update — surprise is the signal.

19. B ***3. Study that raises the difficulty***

The tell is usually that you can state the result and not the condition. The mean of the sample estimates the population mean is correct, and it tells nobody whether you know what randomisation is doing there. That state — fluent, correct and hollow — is common and invisible on paper. The question that settles it is to ask for three questions that would distinguish somebody who understands this from somebody who has memorised the standard statement.

20. A ***3. Study that raises the difficulty***

Matching instruction to a visual or auditory preference has been tested and does not improve outcomes: believe your preferences about what is pleasant, not about what works. Highlighting is consistently rated among the least effective techniques, because it produces recognition rather than recall, and re-copying notes is copying rather than encoding. Rewriting notes from memory is a completely different activity and is excellent.

21. C ***3. Study that raises the difficulty***

Working through gaps with an assistant is engaging, the answers keep coming, and forty minutes disappear, so you finish with a good understanding of one gap and no retrieval practice at all. Set a timer for that block specifically and stop when it goes, even mid-problem. An unfinished problem is a better start to the next session than a finished one, because you begin with retrieval already in progress.

22. B ***4. Subject by subject***

Generated quotations from literary works are usually near-misses: the right sense, the wrong words, or words assembled from different parts of the text. In a subject where the argument is built on exact wording, that is fatal and immediately visible to anybody who knows the work. Every quotation gets checked against the text in front of you, with a line or page reference — not against a search result. Project Gutenberg and the Internet Archive are free and searchable.

23. C ***4. Subject by subject***

Reading the consensus first means you will notice what you were told to notice, and the resulting essay reads exactly like what it is. Afterwards, the same information is genuinely useful, because it tells you whether your observation is original, obvious or contested. The strongest thing in a literature essay remains a small, precise, surprising observation about a few words on the page, defended — and it depends on a reading nobody else has done.

24. B ***4. Subject by subject***

Fossilisation is an error that becomes permanent because you have produced it often enough without correction that it now feels right, and it happens to learners who produce a great deal without feedback. Assistants are tuned to be agreeable and will frequently understand and carry on rather than flag a mistake, which gives you a pleasant fluent conversation full of errors nobody mentioned. The fix is one clause, used every time: list my errors and the correct form, then continue.

25. D ***4. Subject by subject***

Independence is the assumption that quietly invalidates analyses: repeated measurements on the same person, students within classes, sites within regions. If observations are clustered, an analysis that ignores the clustering is wrong regardless of what the output says. Normality is a condition on the residuals rather than the raw data, and a significance test for it is oversensitive at large n and underpowered at small n, which is the opposite of useful — check with a plot.

26. C ***4. Subject by subject***

Generated theses tend towards the balanced and the safe — the causes were both economic and political, the answer depends on context — which is true, unarguable and unmarkable, because a marker cannot reward a position nobody would dispute. The other three failures travel with it: generic evidence without page numbers, an objection stated weakly enough to dismiss for free, and paragraphs that could be reordered without loss.

27. B ***4. Subject by subject***

A rule from one country's statute presented as a general principle is wrong for you in a way that is invisible from the text. Alongside that, fabricated case citations are the highest-profile AI failure in the professional world, and lawyers in several countries have been sanctioned for filing briefs containing invented cases with plausible names, reporters and paragraph numbers. Free primary sources exist: Indian Kanoon, BAILII, CourtListener, AustLII, CanLII, and official legislation sites.

28. A ***4. Subject by subject***

A marker can see in one glance whether the person understood their measurement or copied a number, and models routinely give ten decimal places or round to two when the data supported four. The rules are short and yours to apply: multiplication and division carry the significant figures of the least precise input, addition and subtraction the decimal places, exact counts do not limit precision, and you round once at the end rather than at every step.

29. C ***5. Research, sources and evidence***

Is this real is a question about the world, and the model answers it from the same process that produced the citation, so it will usually say yes — and will occasionally say no about a real paper, which is more confusing still. There is no database being consulted and no lookup being fumbled. The only checks that work are outside the conversation: resolve the DOI, search the exact title, and open the thing.

30. B ***5. Research, sources and evidence***

A page number forces you to say what the source actually shows rather than what it is generally about, and it is what makes the citation yours rather than something you inherited. It is also what lets you answer, out loud before submission, what each reference argues and which page you used. Any reference where the answer to that is no comes out of the list — not gets checked later, comes out.

31. C ***5. Research, sources and evidence***

The limitations section is where honest authors tell you what the paper cannot support, and it is exactly what you need in order to cite it accurately. Writing the four lines in your own words matters for a second reason too: a copied summary is a copying-into-the-essay accident waiting to happen, and paraphrasing is where you find out whether you understood. Thirty papers read this way is a working knowledge of a literature; thirty AI summaries is a folder.

32. C ***5. Research, sources and evidence***

Review filters out a lot of obvious rubbish, usually on a few unpaid hours of work. It does not verify the data, does not usually recompute the analysis, and does not guarantee replication — the replication crisis in psychology and several other fields is exactly this. Peer reviewed is a floor, not a certificate. Preprints have not been reviewed and are legitimate to cite in many fields provided you say so, which is normal practice rather than an admission.

33. D ***5. Research, sources and evidence***

It is an excellent starting point and not a citable source, and both halves hold for the same reason: it summarises, and it tells you what it summarised from. Go to those sources, read them, cite them. The talk page is unusually informative as well, since it is where disputes about the content are argued out, so you can see at once whether a claim is contested. Kiwix packages it for offline reading, free, which matters with intermittent data.

34. C ***5. Research, sources and evidence***

Lateral reading is the technique that research on this consistently finds effective: instead of scrutinising the page in front of you, leave it and find out who is behind it. Students left to themselves stay on the page and evaluate its design, which is precisely what a designed page is for. Everything else that still works is about provenance rather than appearance — who published it first, what the date is, and whether a primary record exists.

35. B ***5. Research, sources and evidence***

These modes are genuinely more thorough, and they concentrate the risk rather than removing it, because a long report with sixty citations invites acceptance precisely because checking it looks expensive. Use it to find out what to search for, then do your own reading of the sources that matter. A bibliography assembled by a tool you did not verify is a bibliography of sources you have not read, which is the misconduct finding rather than a shortcut around it.

36. A ***6. Writing, and keeping your voice***

This is the honest and most common cause, and no technique substitutes for it: go and read three more things and take notes. The others have their own fixes — freewriting for an undecided position, a capitalised placeholder for a first sentence that has to be perfect, and twenty-five minutes on a timer with the internet off for avoidance. Diagnosing which one you have takes thirty seconds and saves an evening.

37. C ***6. Writing, and keeping your voice***

A thesis is fifteen words and it is the entire piece, so the shortness of the output is misleading. If you have already asked and are looking at one, do not simply use it and do not simply discard it: write down two other positions somebody could take, then choose between all three and write out the reason. That converts a supplied answer back into a decision, which is what the stage was for. The other three requests are interrogative and leave the position with you.

38. D ***6. Writing, and keeping your voice***

The commonest failure is a paragraph that presents evidence and then moves on, leaving the reader to work out why it mattered — and a marker will not do that work for you. Analysis is usually the largest single band in a rubric. A quick check on your own draft is to highlight the analysis sentences: a paragraph with none is description. It is also the first thing students cut when over the word limit, which drops the essay a band.

39. D ***6. Writing, and keeping your voice***

Cut throat-clearing, background your reader already has, long quotations that reduce to a phrase, and any paragraph whose removal you cannot immediately object to. What you never cut is analysis, the objection and the specificity of your evidence. Students under a limit tend to cut exactly those three, because they are the hard parts and they look expendable, and the essay drops a band for it.

40. A ***6. Writing, and keeping your voice***

Academic writing means precise, and precision often means the plain word. What the register actually requires is hedging claims to the strength of the evidence, defining terms used in a specific sense, attributing positions to whoever holds them, and being explicit about how sentences relate — none of which needs utilise or heretofore. Usually the model raised the vocabulary and not the precision, and inflated vocabulary is easier to see through than plain prose.

41. A ***6. Writing, and keeping your voice***

Most students read the grade and close the file. Collect the comments from several pieces into one document and look for what repeats: almost everybody has two or three persistent faults — not enough analysis, unsupported claims, does not answer the question — and fixing one raises every future mark. Pasting the comments and asking what patterns appear is analysis of supplied text, the safe category, and it surfaces things you have stopped seeing.

42. A ***6. Writing, and keeping your voice***

What is assessed is whether you can explain something, in order, to people, and answer questions about it; the slides support that. If they could be read without you, you have written a document and are now reading it aloud. Write the talk first and let the slides follow: one idea per slide, the heading carrying the point rather than the topic, six words is plenty, and anything you would read aloud does not go on the slide at all.

43. B ***7. Rules, honesty and the record***

The most specific rule wins, and course-level rules override general ones and are frequently stricter. Look in order: the module handbook, the brief for this piece of work, the integrity policy, then any separate generative-AI guidance. Where they are silent, ask in writing and keep the reply — a dated email from the person marking your work is worth more than any argument you can construct later, and email rather than a corridor conversation, because honest people misremember.

44. D ***7. Rules, honesty and the record***

Several policies make undisclosed use the offence rather than the use, so something entirely permitted becomes misconduct simply by not being declared, which is a trap for the honest. There is no penalty anywhere for declaring a permitted use and a real one for omitting a use that mattered, so over-disclosure costs two extra sentences. Write it as you go, in your working file, so the statement is accurate rather than reconstructed the night before.

45. A ***7. Rules, honesty and the record***

Declaring that you generated the essay does not turn generating the essay into acceptable practice under a policy that forbids it; it makes the finding faster. A disclosure also does not cover what it does not mention, and one that omits a use you made is worse than none, because it is now an inaccurate statement rather than an absent one. If a use is uncomfortable to write down, that discomfort is the information, and the time to act on it is before submission.

46. A ***7. Rules, honesty and the record***

An allegation is not a finding, procedures exist, and people are cleared regularly. The single most damaging thing students do is admit to something they did not do, because the process is frightening and admitting looks like the fastest way out — a finding is lasting and far harder to undo than to contest. Equally, do not lie: if you did do something, early admission is treated more leniently everywhere. The other three options are all correct moves.

47. B ***7. Rules, honesty and the record***

A dyslexic student using text-to-speech still has to have read and understood the material; a student dictating still has to have the argument. None of those substitutes for the thinking, and all of them remove a barrier that was never the subject of the assessment. Register it with the disability service, because a documented adjustment is a complete answer if a question is ever raised, rather than an explanation offered afterwards.

48. D ***7. Rules, honesty and the record***

These are three separate questions with different answers, and collapsing them into one worry is what makes them unanswerable. Almost everything is stored for some period regardless, human review happens in limited circumstances such as abuse investigations and legal process, and the training toggle is usually yours: consumer free tiers commonly train on input by default, while business and education tiers commonly contract not to. Find it once, per tool, decide, and stop thinking about it.

49. C ***7. Rules, honesty and the record***

The contract usually prohibits training on your input and places the data under an agreement the institution has already assessed, and it is often a paid tier at no cost to you, so for coursework it is generally the right choice. The cost is that it is a work account: administrators are not reading conversations for entertainment, and records may be accessed under the institution's own procedures. That is a reason not to treat any account, personal or institutional, as a diary.

50. B ***8. Assessment, and the years after it***

You get somebody else's compression of material you needed to compress yourself, presented with maximum clarity, which produces the maximum feeling of understanding for the minimum retrieval. If you want the summary, write it from memory first and then ask the model what you left out — that sequence is one of the best exercises in the course. The other three options are all good uses of the same hour.

51. D ***8. Assessment, and the years after it***

Feelings about which topics are weak are generated by the same fluency illusion that produced the gap, so they are exactly the wrong instrument. One line per lookup, no self-criticism, recorded at the moment you did not know — at revision time that list is your syllabus, and it is honest in a way no later reconstruction can be. The confidently-wrong list and the unaided-check log work for the same reason.

52. A ***8. Assessment, and the years after it***

Describe, explain, evaluate, justify and compare have fixed meanings in most examination systems, and answering a different one costs the whole question however good the material is. The rest of a mark scheme is just as informative and almost nobody reads it: what earns the marks is frequently a specific term, a specific comparison or the statement of an assumption, and the difference between the top two bands is usually one identifiable thing.

53. B ***8. Assessment, and the years after it***

It does three things at once: it converts reading into thinking rather than leaving it to a panic at the end, it means you always have text so the blank page never arrives, and it produces a continuous timestamped record of the work developing, which is the process record in its strongest form. Students who write only at the end lose most of what they read in months one and two, because they never encoded it.

54. A ***8. Assessment, and the years after it***

A generated statement is recognisable not because of any detector but because it says the same things: passion from a young age, a formative moment described in general terms, transferable skills, a closing line about contributing. That is the average of what has been written, which is precisely what does not distinguish you in a document whose only purpose is to distinguish you. What survives the test is short, specific and unfakeable.

55. B ***8. Assessment, and the years after it***

You are governed simultaneously by the university's academic policy, the host organisation's rules and often a professional body's code; they do not say the same things, and the strictest applies. Students get into difficulty here not through dishonesty but by importing a university rule into a setting where a stricter one governs. Ask in the first week, in writing, and where there is genuinely no policy the safe default is that nothing belonging to the organisation goes into any external tool.

56. B ***8. Assessment, and the years after it***

Both answers are legitimate on different days, and some weeks are document weeks. But if the answer is the document every time, for a year, then at the end of that year you will have a great many documents and will not be able to do anything you could not do at the start. That is the whole risk, it is entirely within your control, and the question takes five seconds at the end of any session.