

Autonomous AI and Education

Michael Gannotti

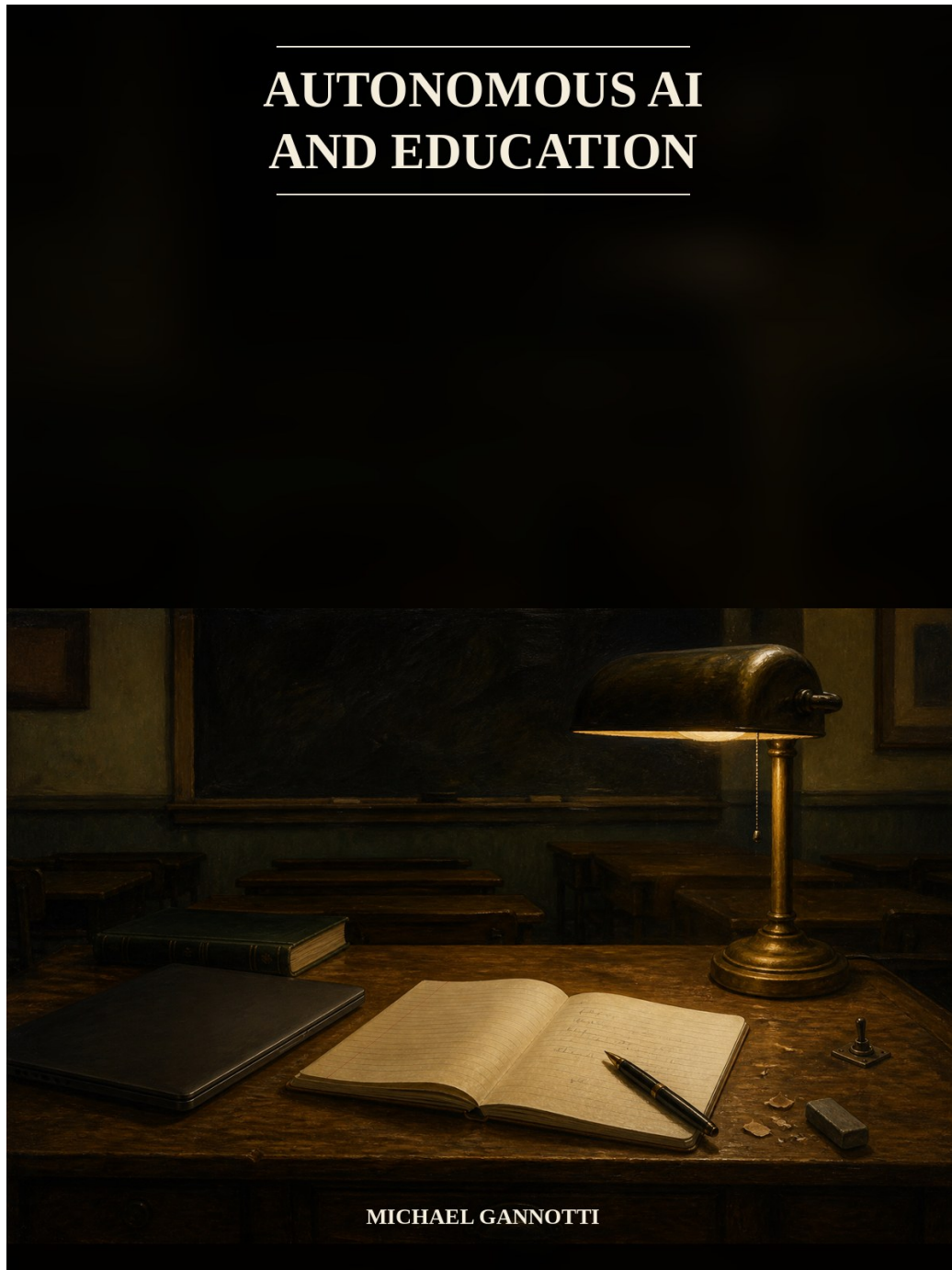


Figure 1: Cover

Autonomous AI and Education

Michael Gannotti

This illustrated edition is the teaching-manual of 30 August 2026. Endnotes run in one series. Chapter-opening images are original still-lives, not portraits of persons or children. Documents still constrain the facts.

Contents

- Welcome
- How to Use This Book This Week
- If You Only Remember Five Things
- Formation Hour and Performance Hour
- Chapter 1 — The Job the Student Still Does
- Chapter 2 — Four Kinds: Chat, Tutor, Adaptive System, Agent
- Chapter 3 — The Exam After the Window Closes
- Chapter 4 — Age: A Tool Fine at 17 Can Be Wrong at 8
- Chapter 5 — Struggle Before Rescue
- Chapter 6 — You Cannot Judge What You Do Not Know
- Chapter 7 — Who Hosts, Who Turns It Off
- Chapter 8 — What “Done” Looks Like: Two Lanes
- Chapter 9 — Records a Stranger Can Read
- Chapter 10 — Resources, Named by Fit
- A Note on Sources
- Bibliography

Welcome

This book exists because the hour at your table is still the program. There is no department down the hall that will form the skill after the window closes. There is you, a student, and a tool that can finish a page in seconds. That last fact does not retire the hour. It makes the hour the work.

I wrote this for a capable, busy, willing parent — and for a teacher sitting with one child or a room. You may be teaching two ages at once. You may be fitting school between a job, a toddler, and a grocery run. You may already use a chatbot for your own mail, or you may still be deciding whether any of this belongs near a nine-year-old. Many adults feel behind the product names. That feeling is common. It is not a verdict. We will move on from it. This book will make you fluent enough in this week's idea to notice a wrong turn and ask a good question — without handing the thinking to a machine.

You do not need to be a computer scientist. You do need to understand today's idea well enough to hear “the homework looks perfect, so they learned” as a mix-up of two different things, not as a compliment. You need a session shape you can run on a Tuesday. You need a few sentences that actually help. That is the job. A path through this subject — from what the student must still do, through the four kinds of tool, through the exam after the chat window closes, through age, struggle, judgement, and who hosts the machine — is possible at a kitchen table. The path is not a personality trait and it is not a percentile. It is a sequence of ideas, practiced until they hold, with you in the chair.

About one in four U.S. teens already used ChatGPT for schoolwork in late 2024, roughly double the year before.¹ That figure is context, not a ranking and not a

¹Olivia Sidoti and Jeffrey Gottfried, “About a Quarter of U.S. Teens Have Used ChatGPT for Schoolwork — Double the Share in 2023,” Pew Research Center, 15 January 2025. Among U.S. teens ages 13–17, 26 percent said they had used ChatGPT for schoolwork, up from 13 per-

report of agents. You are one of those tables, or you soon will be. The hour in front of you is the work.

What a good hour looks like

There are two hours in this book. Learn the difference once. Later chapters will not rebuild it.

A *formation hour* is for a skill that must still live in the student when the tool is taken away: retrieving a fact, writing a first draft, placing a fraction, tracing a proof, noticing an error. The child attempts first. Help, if any, is a hint. The check at the end is unaided. Book closed. No chatbot in the chair.

A *performance hour* is different, and it comes later, once the student can already do the core move and can catch a fluent wrong answer. Then a tool may sit on the table as a partner — logged, named, and judged. What you grade is the judgement: what they kept, what they cut, what they checked, and whether they can defend it out loud.

You sit down already knowing which hour it is. The child warms up on something they can already get right. You model one move out loud, short. Then they try, and you wait. You ask one good question — not “did you get it?” but “how did you know?” Practice mixes the new move with last week’s. The hour ends with one or two problems they do alone. You stop talking sooner than feels polite. The child holds the pencil. You hold the key. That is the hour. The next long piece of this front matter, *Formation Hour and Performance Hour*, will teach it in full.

What you will actually get

Each teaching chapter does eight jobs, always in the same order, so you are never hunting for the move.

You will learn why this week’s idea is worth the struggle — what it unlocks later. You will understand the idea yourself, in plain language, with one everyday picture and one precise picture, and with the wrong answers you should be able to hear.

cent in 2023. Chat use, not agent use. <https://www.pewresearch.org/short-reads/2025/01/15/about-a-quarter-of-us-teens-have-used-chatgpt-for-schoolwork-double-the-share-in-2023/>. Access date for URLs in these notes: 30 August 2026.

You will get a session you can run this week, including exact wording. You will get practice that builds learning, not a pile of the same problem. Your student will get a short section of their own. If it isn't clicking, you will get three diagnostics and a next move for each, with no shame in the room. Tools, including AI, stay optional and adult-side; the rules live once, in the session chapter. And you will get a plain checklist for "done enough," so you can place by skill rather than by birthday.

This week you can learn the two hours and today's idea well enough to hear a wrong answer. Today the child can try one problem without a machine holding the pencil.

What this book will not do

This book will not hand you 180 days of worksheets, prompts, or "AI lesson plans." A year of photocopies is not a teaching method, and I will not pretend it is.

It will not sell you a product. Later, a short resources chapter names common tools by fit — parent load, age band, how much the student still has to think — so you can choose. A chatbot, a hint tutor, an adaptive practice program, and an agent that keeps working after you leave are different objects. None of them is this book, and this book is not a catalog in disguise.

It will not promise a score. There is no guaranteed percentile, no guaranteed SAT number, and no guaranteed college letter inside these pages. What this book can promise is a path: the ideas in order, at the skill the child actually has, until they can do the next idea unaided — and, later, until they can use a tool without losing the skill.

It is not a ban tract and it is not a cheer. A tool that finishes the page can still be useful on the adult side of the table. A tool that finishes the page during the attempt can starve the very skill the hour was for. Both can be true in the same kitchen.

And this is not a book that lectures your child about you, or you about your character. The student is a person, not a percentile. When this book speaks to them, it speaks with respect. No baby-talk. No research sermon.

The promise

If you only remember one sentence, remember this: the student still does the thinking.

School exists to form skills that are still required when the tool is taken away: retrieving, generating, catching errors, knowing enough about a subject to notice a fluent lie, and the habit of checking. An AI world does not retire those. It makes them the scarce resource.

You can do this. You do not have to know next year's product today. You have to know this week's idea well enough to sit still while they struggle, then ask one good question. Start here. Read *How to Use This Book This Week*, then the one-page list of five things, then *Formation Hour and Performance Hour*. After that, open the chapter that matches the skill in front of you. You will know more after one chapter than you know this morning. Your child will have something to try today. We can do this.

How to Use This Book This Week

Start at the skill in front of you, not on page one because a catalog, a birthday, or a well-meant relative said so.

This book is a path, not a calendar. The teaching chapters run: the job the student still does; the four kinds of tool; the exam after the chat window closes; age; struggle before rescue; judgement; who hosts the machine; then what “done enough” looks like across both hours. Records and resources come last. Chapter labels here are sequence anchors. They tell you where an idea usually sits. They do not tell you where *this* child sits.

Pick the chapter by skill, not by birthday

Open the chapter you think is right. Skip to **What “done enough” looks like** at the end of the *previous* chapter, or to **If it isn’t clicking** in the one you opened. If the child can already do those checks unaided, you are too early. If the checks from two chapters back are still failing, drop back.

If they still treat every fluent paragraph as true, stay in Chapter 1 even if they are sixteen. If they cannot tell a chatbot from an agent, stay in Chapter 2 even if the laptop already has both. If helped practice always looks fine and the unaided check always drops, stay in Chapter 3. If the tool on the table is built for a seventeen-year-old and the child is eight, open Chapter 4 before you open anything else.

A ten-year-old can sit in three different “grades” of this subject on the same morning: literacy *about* AI, no student-facing chat, and a parent using a tool after the child is in bed. That is ordinary. Age in this book is a design spec, not a birthday present. UNESCO’s guidance puts independent classroom use of generative tools

at thirteen; U.S. children's-privacy law draws a bright line under thirteen.² Both are floors. They are not a method. Use this book's checklists, then teach.

Two children may need two chapters. That is ordinary. The session shape in *Formation Hour* and *Performance Hour* still holds; the idea on the table changes.

How you use the parent half

Most of each teaching chapter is for you. Read it *before* the lesson, not over the child's shoulder.

Why this matters tells you what this idea unlocks. You will teach with more patience when you can see the later door.

For the parent: understand it yourself gives one everyday picture, one precise picture, and three to five wrong answers you should be able to hear. Sit with those. If "the essay looks great, so they learned" still sounds like success, that is the performance-and-learning idea, and it is worth five quiet minutes now.

How to teach it this week assumes the session shape from *Formation Hour* and *Performance Hour*. It will not rebuild the hour. It will give you the first problem, the wording for *this* idea, and how to fade your help.

Practice that actually builds learning is what goes on the page after the model: a short blocked set of the new move, then mixed problems so the child has to choose, plus a little retrieval of facts they already know. Kitchen talk and news stories can motivate or give a second picture. They do not replace the unaided problem.

Tools, including AI is optional, short, and for the adult. The rules live once in *Formation Hour* and *Performance Hour*.

What "done enough" looks like is how you leave. You are looking for unaided work, not a perfect Tuesday.

²UNESCO, *Guidance for Generative AI in Education and Research* (Paris: UNESCO, 2023): independent classroom use of generative AI recommended from age 13; use of untested tools with children under 13 described as not appropriate. Children's Online Privacy Protection Act, 15 U.S.C. §§ 6501 *et seq.*, and the COPPA Rule, 16 C.F.R. Part 312: bright line under 13. The 2025 COPPA amendments (Federal Register, 22 April 2025) set a general compliance date of 22 April 2026. Access date for URLs in these notes: 30 August 2026.

You do not have to read the whole chapter tonight. You do have to read the parent half of *this week's* idea before you sit down with the child.

The five-minute parent warm-up

Make this a habit.

Five minutes. Child not yet in the chair. Phone face down.

1. Read today's idea until you can say it in one sentence.
2. Work two problems yourself, on paper, without the key until you finish. In this book a "problem" may be: name the skill that must remain; classify a tool; close the window and do a twin item; match a tool to an age band.
3. Glance at the "wrong answers you should be able to hear." Name the one you would have given last year.
4. Write one sentence you will actually say. Not a speech. Example: "The tool can finish the page. You still do the thinking." Or: "A chatbot waits. An agent keeps working after you leave."
5. Close the book to the student page. You are ready.

That five minutes is how you stay out of their working memory. If you are learning the idea *while* they are stuck, you will talk too much or hand them the machine. Prepare first. Then sit still. You need today's sentence, not a seminar.

How the student uses "For the student"

Every teaching chapter includes a short section written to the student, not about them. One or two pages. Warmer. Direct. What the idea is, a tiny worked example, two tries, an "explain it back" prompt, and one challenge problem.

Hand them that section when it is time for their attempt — after your short model, not instead of it. They may read it silently or out loud. You stay in the room. You do not add a lecture about college.

The two tries are theirs. "Explain it back" checks that they have a picture, not only a copied step. The challenge is optional that day; it is there so a ready student does not sit in busywork.

The student page is not something to send off with an unsupervised chatbot. The attempt is still theirs. You still hold the key.

When to skip ahead

Skip ahead when this chapter's "done enough" checklist is already true *unaided*. Lingering past that is not thoroughness. It is boredom wearing a virtue.

Slow down when the same wrong answer repeats after a clear model and a real attempt. **If it isn't clicking** will give you three likely causes and a next move: a missing picture, a tool in the chair during the attempt, or a check that never takes the tool away. Missing facts about how a tool works are a fluency problem, not a character problem. Ten minutes of struggle on a well-posed new idea is ordinary work.

Diagnose "chat versus agent," "helped practice versus unaided exam," and "this age, this tool" before you buy a new subscription. Accelerate a child who has the pictures. Hold still for a child who does not. Both are the same kindness.

If you need a human tutor, that is a normal high-school plan, not a failure of the kitchen.

For this week: pick the chapter by skill. Do the five-minute warm-up. Run the hour as *Formation Hour and Performance Hour* describes it. Let the student page be theirs. Stop talking sooner than you want to. That is how you use the book.

If You Only Remember Five Things

Keep this page. The chapters will add wording, pictures, and practice. They will not replace these.

1. The student still does the thinking. A tool can finish a page. The hour is for the skill that remains when the page is taken away: retrieve, generate, catch an error, know enough about the subject to notice a fluent miss, and check. School exists to form those. An AI world does not retire them. It makes them scarce.³

2. Chat is not an agent. A chatbot waits for a prompt and returns text. An agent is given a goal, uses tools, and can keep working after you leave. Tool use is not autonomy. Hermes 4 is a language model — an engine. Hermes Agent is a runtime that can run jobs on a machine. They are not the same object. Grok the chat assistant, Grok Bot the teammate with its own computer, and the Grok-as-tutor announcement for El Salvador schools are three different things. As of this writing there is no K–12 student-facing open-web school agent as a product you can buy for a desk.⁴

3. Helped practice is not the exam. Tonight’s fluent page, with a window open,

³Nicholas C. Soderstrom and Robert A. Bjork, “Learning Versus Performance: An Integrative Review,” *Perspectives on Psychological Science* 10, no. 2 (2015): 176–99: current performance is often an unreliable index of learning. Evan F. Risko and Sam J. Gilbert, “Cognitive Offloading,” *Trends in Cognitive Sciences* 20, no. 9 (2016): 676–88: offloading is an old mechanism; generative tools extend it from memory and calculation to drafting and evaluation.

⁴Product documentation for Hermes Agent (hermes-agent.org), Grok Bot (x.ai, 11 August 2026), Cursor Cloud Agents, Claude Cowork, and OpenAI Workspace Agents: goal, tools, unattended runtime. Nous Research: Hermes 4 is an open-weight model family (Teknium et al., arXiv 2508.18255, 25 August 2025); Nous Portal docs state Hermes 4 is not recommended inside Hermes Agent. xAI El Salvador partnership announced 11 December 2025 (Grok as tutor across more than 5,000 schools); no outcome study found as of 30 August 2026. No documented K–12 student-facing autonomous open-web agent program in the research this book stands on.

is performance. Learning is what they can still do when the window is closed. Unguarded chatbot help can raise practice scores and then cut the unaided exam. A hint tutor that withholds the full solution is a different treatment. Students often do not notice the loss. Grade the unaided check.⁵

4. A tool that is fine at 17 can be wrong at 8. Age is a design spec, not a birthday present. Under thirteen, U.S. children’s-privacy law is in force; independent generative chat is not a primary-grade method. Companion chat is not a tutor. Match the tool to the child in the chair, then to the law.

5. The child attempts first. You hold the key. AI is optional for the adult. The student does the problem. You keep the answers. A tool may explain today’s idea *to you*, write extra practice with the key hidden, draft a hint script, or help you diagnose work already done. It may not sit in the chair during the attempt, and it may not paste the worksheet. The full rules are in *Formation Hour and Performance Hour*, once. Later chapters will point back rather than preach.

If this week needs a compass, this is it. Five things. Then sit down and teach.

⁵Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman, “Generative AI without Guardrails Can Harm Learning: Evidence from High School Mathematics,” *Proceedings of the National Academy of Sciences* 122, no. 26 (2025): e2422633122. Unguarded GPT-4-style help raised assisted practice about 48 percent relative to control, then cut unaided exam grades about 17 percent. A hint tutor that withheld full solutions raised practice without that exam drop. A useful study, not a promise that every home will see the same result. Access date for URLs in these notes: 30 August 2026.

Formation Hour and Performance Hour

The hour has a shape. Learn it once. Later chapters will give you today's idea, today's first problem, and today's sentence. They will not rebuild this hour. When a chapter says "run the session," it means this.

You do not need a school bell. You need a beginning, a short model, a real attempt, one good question, mixed practice, and an unaided check. Younger children may finish in twenty-five minutes. Older students may need forty-five. The shape does not change. The idea on the table does.

Sit down having already done the five-minute parent warm-up from *How to Use This Book This Week*. You know today's sentence. The child has the pencil. You have the key.

You also know which hour it is.

Which hour is it?

A *formation hour* is for a skill that must still live in the student when the tool is taken away. New math. A first draft. Tracing a proof. Placing a fraction. Reading a page and saying what it claimed. For this hour, AI is off the table or hint-only. What you grade is the unaided work.

A *performance hour* is for work in a world where tools exist — after the student can already do the core move and can catch a fluent miss. Then a named tool may sit on the table. What you grade is the judgement: what they asked it, what they kept, what they cut, what they checked, and whether they can defend it with the window closed.

Most Tuesdays in childhood are formation hours. Most first encounters with a new idea at any age are formation hours. Performance hours are earned. They are not a reward for finishing fast. They are a different assignment, with a different check.

If you are unsure, it is a formation hour. You can always add a tool later. You cannot un-copy a page.

Nicholas Soderstrom and Robert Bjork's review is the plain fact underneath both hours: what looks fluent *now*, with help in reach, is often a poor index of what will still be there later, alone.⁶ The unaided exit ticket is the check this book trusts.

1. Warm-up

Three to five minutes. Facts and moves they can already get right. Mixed, not a page of the same item. This is retrieval, not a test of character.

Say:

"We're going to start with things you already know."

"Tell me, in one sentence, what a chatbot does."

"What's 7×8 ? Take your time."

If they already get these right untimed, you may use a brief timer. Say:

"We'll time this only because you already get these right. Ready?"

If they do not already get them right, skip the timer. Time only what they can already produce.

Keep this short. The warm-up is not the lesson. When you feel yourself teaching a new idea here, stop. That idea belongs in the model.

2. Short model

Five to eight minutes. One fully worked example, out loud. Then a second example with holes they fill. Then you stop.

⁶Nicholas C. Soderstrom and Robert A. Bjork, "Learning Versus Performance: An Integrative Review," *Perspectives on Psychological Science* 10, no. 2 (2015): 176–99. Current performance is "often an unreliable index" of learning. The unaided exit ticket is the check this hour trusts.

New, dense material — first time naming the skill that must remain, first time telling chat from agent, first time closing the window and doing a twin problem — is where a short model earns its keep. A novice staring at a blank page spends working memory searching, not learning. Show the move. Then fade.⁷

Say:

“Watch this one. I’ll say every step out loud.”

“The tool can finish the page. You still do the thinking. That’s why I close the window before the check.”

“What is this hour for?”

Then, on the second example:

“Your turn to tell me what I write next. I’ll hold the pencil this time. You do the thinking.”

“I’m leaving the last step blank. You finish it.”

Include one incorrect example when the idea is new, and ask:

“What did they think the tool was doing?”

That question is worth more than another correct copy.

When they can already execute the move, stop showing full examples. Extra modeling after they can do it is noise. Fade: full example, completion problem, conventional problem. Then they work.

The model is short on purpose. If you are still talking at minute twelve, you are giving a lecture. Close the model. Hand them the first problem.

3. Student attempt

This is the center of the hour. Eight to twelve minutes. Maybe longer for a multi-step task. The first problem is theirs.

⁷John Sweller and Graham Cooper, “The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra,” *Cognition and Instruction* 2, no. 1 (1985): 59–89: for novices, a worked solution outperforms unguided search. Alexander Renkl and Richard K. Atkinson developed guidance fading (full example, then completion problems, then conventional problems). Once the learner can execute the procedure, extra full examples become redundant.

Hand them “For the student” if the chapter has it, or the first problem from **How to teach it this week**. Then you talk less than you want to.

Say:

“This one is yours. I’ll be quiet.”

Then be quiet.

If they stall, use this order: ask, wait, hint, then model. Not the reverse.

Ask:

“Show me what you tried.”

“What do you already know about this?”

“Is this the student’s job, or the tool’s job?”

Wait. Count a slow ten in your head. The silence is the work. If you fill it, you took the problem back.

Hint, one hint:

“Look at what would still have to be true if we closed the laptop.”

“Name the kind of tool first. Then decide.”

“You already know a fact that would help. Which one?”

Then, if they are still stuck after a real try:

“I’m going to show you this one step. Then you take it from here.”

Show the step. Return the pencil.

If they are missing a fact — they cannot say what a chatbot does, they cannot tell helped work from unaided work — that is a hole, not useful struggle. Fill the hole with a short retrieval or a picture, then return to the problem. Struggle on a well-posed new idea, with the pieces in place, is the lesson. Struggle on missing pictures is just being lost.

Keep your hands off their pencil, and keep the chatbot off the desk during a formation attempt. Taking the pencil feels kind. Opening the window feels kind. Both teach them to watch. If you feel the urge, put your hands in your lap and say the ask-sentence again.

4. One good question

Two or three minutes. Not “did you get it?” They will say yes. You will learn nothing.

A good question makes them generate the idea again, in words or in a second representation.

Say:

“How did you know?”

“Can you show me a second way?”

“What would happen if we closed the window?”

“Explain it so I could do the next one.”

“Whose job was that step?”

One question. Maybe a follow-up. Then stop. This is not a seminar. You are checking that they have a picture, not collecting a speech.

If they cannot explain it back, they copied a step. Return to a picture — the two hours, the four kinds, the unaided twin — and give one more attempt. Copying is what a stuck novice does when the model was too fast or the attempt was too short. Slow the model tomorrow. Keep today’s pencil theirs.

5. Mixed practice

Ten to fifteen minutes. First a short blocked set of the new move — four to six problems that use today’s idea — so they learn to execute. Then mix. Yesterday’s type. Last month’s type. A word problem of a known kind. The new move sitting next to an old one so they have to choose.

Blocked practice teaches execution. Mixed practice teaches *when* to use the move. A page of twenty identical problems feels like mastery at 9 p.m. It often is not still there in two weeks.⁸

⁸Doug Rohrer, Robert F. Dedrick, and Kaleena Burgess, “The Benefit of Interleaved Mathematics Practice Is Not Limited to Superficially Similar Kinds of Problems,” *Psychonomic Bulletin & Review* 21 (2014): 1323–30. After a short blocked introduction of a new skill, mix yesterday’s and last month’s types. The same shape holds when the “skill” is classifying a tool or closing a

Say:

“These first few are today’s move. Get the feel of it.”

Then:

“These are mixed on purpose. First decide what kind of problem it is. Then choose.”

“You already know this type from last week. Find it among the new ones.”

Kitchen, news, and “a friend used ChatGPT for the essay” can motivate, or give a second picture of a move they are already practicing on paper. They do not replace this page. A child who can talk about AI at dinner and cannot close the window for a twin problem has performed, not learned.

Keep practice short enough that attention holds. More pages are not more learning if the last ten are sloppy. Stop while they can still think.

6. Unaided exit ticket

Three to five minutes. One or two problems. Book closed. No hints. No “For the student.” No chatbot. No looking at the worked example.

This is the check. Tonight’s helped practice, with you in the chair and a tool in reach, is not this check.

Say:

“Two problems. Book closed. I won’t help. That’s the point.”

“Wrong is useful. We’ll look after.”

“You don’t need a perfect score. I need to see what you can do alone.”

Then you are quiet again. When they finish, you look. If both are right, you are done enough for today. If one is wrong, you diagnose — which wrong answer from the parent section is this? — and you note tomorrow’s first move. You do not redo the whole hour. You do not punish. You find the picture that was missing.

An exit ticket that always needs you, or always needs a window, is not an exit ticket. Fade until it doesn’t.

window.

When to stop talking

After you ask, wait. If you fill the silence, you took the problem back.

During the attempt, your job is almost nothing. Presence, not narration. If you hear yourself explaining while they write, you have moved back into the model. Stop. Let them write.

Instead of “this is the hard part” or “I was never good at computers,” say today’s sentence: “Let’s look at whose job that is.” The first sentence fills their working memory with your feeling. The second sentence teaches.

Praise a clear picture, a second method, a catch of their own error — not speed.

“You noticed the fluent paragraph still had to be checked. That’s the judgement I want.”

Stop talking when the next sentence would be a rescue. Hands in your lap. Count ten. Then one hint, or one step, and the pencil goes back.

If you remember only one move from this page, remember that one.

The AI rules, once

Later chapters will point here in one sentence. They will not reprint this box.

Child first. You hold the key. AI is optional for the adult.

The student attempts the problem before any tool is in the room. You keep the answer key — on paper, in a closed tab, in your head after the warm-up. A chatbot does not sit beside them during a formation attempt.

Use a tool, if you use one at all, for four jobs: explain today’s idea *to you* before the lesson; make extra practice with the answers hidden from the child; write a hint script you can read from; diagnose work they already produced.

Never paste the worksheet, ask “what’s the answer,” photograph the page for a key, or leave an unsupervised chatbot in their hands during the attempt. Hamsa Bastani and colleagues found that an unguarded

chatbot raised practice scores and then cut the unaided exam. A hint tutor that withheld the full solution was a different treatment.⁹

Facts are not a language-model job. Use a tool to generate a short oral list or a practice deck. The child retrieves the fact.

You can teach this entire book with no AI. Many families will. The hour still has the same shape.

If a product claims it “does 100 percent of the teaching,” treat that as advertising. You still need to hear a wrong answer. You still hold the key.

A chatbot waits for a prompt. An agent takes a goal, uses tools, and can keep working after you leave. This hour is about the student, not about buying the newest runtime. There is no K–12 student-facing open-web school agent as a product for the desk. What schools have actually bought is teacher-side help and constrained tutors. Your kitchen does not need a different rule.

A Tuesday, said plainly

Here is an illustration, not a reported family. A parent has spent five minutes writing the sentence “the tool can finish the page; you still do the thinking.” The child warms up by naming two things they already know how to do without a laptop. The parent models one homework task out loud, circling the steps that must remain in the student, then leaves the last circle blank. The child marks a second task, slowly. The parent asks, “What would still have to be true if we closed the window?” Mixed practice: two tasks to mark, one yesterday fact, one short unaided item. Exit ticket: “This paragraph was written with a chatbot. Close the laptop. Tell me the claim in your own words. I won’t help.” Right or wrong, the hour had a shape. Tomorrow’s first move is already visible.

You can run that hour. You do not need a degree. You need today’s sentence, a pencil you refuse to take, and the willingness to stop talking.

⁹Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman, “Generative AI without Guardrails Can Harm Learning: Evidence from High School Mathematics,” *Proceedings of the National Academy of Sciences* 122, no. 26 (2025): e2422633122. In a field experiment, a ChatGPT-like tool raised assisted practice grades about 48 percent relative to control, then cut unaided exam grades about 17 percent. A hint tutor that withheld full solutions raised practice without that exam drop. Not a homeschool study; the transferable caution is the crutch. Access date for URLs in these notes: 30 August 2026.

Later chapters will say: run the session as in *Formation Hour and Performance Hour*. That sentence is this page. The idea will change. The hour will not.

Chapter 1

The job the student still does



Figure 2: A laptop, pencil, paper, clock, and closed book on an empty desk.

Why this matters

The machine can finish the page. That is new. What is not new is the reason you sat down: a person who can still do the work when the machine is gone.

That is the job of this chapter. Not a product tour. Not a ban. The job the student still does.

A calculator did not retire number sense. A spelling checker did not retire knowing what you meant. A search bar did not retire reading a page and saying what it claimed. Each tool moved some labor off the pencil. Each tool made the remaining labor more important, not less. Generative tools go further. They can draft, explain, plan, and sound sure. They can do the very moves school exists to form: retrieving, generating, catching errors, knowing enough about a subject to notice a fluent miss, and checking.¹⁰

If those moves leave the student, the page still looks finished. The person does not.

That is why this idea is worth the struggle. Later chapters will name four kinds of tool, show you the exam after the window closes, and match tools to age. None of that helps if this week's picture is missing. If "the homework looks perfect" still sounds like learning, you will buy the wrong hour. If "the tool will think for them" still sounds like kindness, you will take the pencil without meaning to.

What you are teaching, as a skill, is a split the child can run on a Tuesday: *this step is mine; that step may later be a tool's*. Retrieval is theirs. A first attempt is theirs. Noticing that a fluent paragraph does not match the source is theirs. Pressing send on a finished page, later, after they can already do those things, may be shared. The split is the idea. The unaided check is how you see it.

Grade labels in this book are sequence anchors, not birthdays. An eight-year-old and a sixteen-year-old may both sit here. The eight-year-old is learning that *they* still count, write, and explain. The sixteen-year-old is learning the same split on a harder page. About one in four U.S. teens already used ChatGPT for schoolwork in late 2024, roughly double the year before.¹¹ That is chat use, not agent use. It

¹⁰Evan F. Risko and Sam J. Gilbert, "Cognitive Offloading," *Trends in Cognitive Sciences* 20, no. 9 (2016): 676–88.

¹¹Olivia Sidoti and Jeffrey Gottfried, "About a Quarter of U.S. Teens Have Used ChatGPT for Schoolwork — Double the Share in 2023," Pew Research Center, 15 January 2025. Ages 13–17; 26 percent versus 13 percent in 2023. Chat use, not agent use.

is still the pattern in the room. Estonia’s AI Leap exists in part because so many students were already using free commercial bots to finish homework faster — the pattern this chapter treats as a crutch unless you design around it.¹² Your kitchen does not need a national program. It needs today’s sentence.

What a child of a given age can learn here is mostly what they have already had a chance to practice, not a lock on a birthday. A twelve-year-old who has never closed the window for a twin item sits in this chapter. A sixteen-year-old who already drafts unaided and then uses a tool to check a citation may only need this chapter as a named split, then move. Both can be right. The checklist at the end tells you, not the catalog.

Paid tools and free windows are not the same object. Among Harvard undergraduates who used AI, about 40 percent of students with no financial aid had paid for an AI product, versus about 20 percent of students on aid.¹³ Awareness of chat is spreading. Access to a stronger, quieter, better-configured tool is not. That gap does not change the job the student still does. It makes the unaided hour more important, not less: the family that cannot buy a premium window should not be told the skill now lives in a subscription. The family that can buy one should not be told the subscription is the skill.

You do not need to be a computer scientist. You do need to understand this week’s idea well enough to hear “the essay looks great, so they learned” as a mix-up of two different things, not as a compliment.

For the parent: understand it yourself

Many adults feel rusty around the product names. That is ordinary. Five minutes of this section, then the warm-up at the end, is enough for tomorrow.

Everyday picture. You can drive a car with a map app speaking every turn. You arrive. If the phone dies on the way home, you may not know the neighborhood. The trip looked fluent. The knowledge was in the speaker, not in you. Homework

¹²Estonia’s AI Leap is a national program responding in part to widespread student use of free commercial chatbots for homework; cited here as context for the pattern, not as a homeschool mandate.

¹³Harvard undergraduate survey, arXiv:2406.00833 (2024): among AI users, 40 percent of students with no financial aid paid for AI products versus 20 percent of students on aid. A paid-tier gap, not an agent-use statistic.

with a chatbot in the chair is that trip. The page arrives. The neighborhood may not.

A second picture, closer to the table: a child uses a calculator for 7×8 before they can produce 56. The answer is right. The fact is not theirs. Later, when a word problem needs 7×8 inside a larger move, they stall. The calculator was not the villain. Timing was. Use it after the fact is theirs. The same timing rule is this chapter.

Precise picture. Psychologists have long separated *performance* from *learning*. Performance is what you can do right now, in this condition, with whatever help is in reach. Learning is a lasting change — what you can still do later, in a new condition, often with the help gone. Conditions that make practice look easy often produce weaker long-term skill than conditions that feel a bit harder: generating an answer, spacing practice, mixing types, retrieving instead of restudying.¹⁴

Cognitive offloading is the ordinary move of putting a piece of thinking onto an object: a list, a calendar, a calculator, a search. It is not a moral failing. It is how humans already live. Generative tools extend the same move from memory and arithmetic to drafting, explaining, and evaluating — the skills the hour was for.¹⁵ When the offloaded process is the learning objective, the page inflates and the person thins.

So the job the student still does, named as moves:

- **Retrieve.** Bring a fact or a step back without looking it up. The capital, the times table, the claim of the paragraph they just read.
- **Generate.** Produce a first attempt: a sentence, a diagram, a plan, a guess that can be wrong.
- **Catch an error.** Notice that $1/2 + 1/3$ is not $2/5$, or that a fluent paragraph cites a date the source does not have.
- **Know enough.** You cannot judge a history paragraph if you do not know the period. Critical thinking is not a bicycle skill you can teach in the abstract while starving the subject.¹⁶

¹⁴Nicholas C. Soderstrom and Robert A. Bjork, “Learning Versus Performance: An Integrative Review,” *Perspectives on Psychological Science* 10, no. 2 (2015): 176–99. Robert A. Bjork and Elizabeth L. Bjork, “Making Things Hard on Yourself, But in a Good Way: Creating Desirable Difficulties to Enhance Learning,” in *Psychology and the Real World*, 2nd ed. (2011).

¹⁵Risko and Gilbert, “Cognitive Offloading” (2016).

¹⁶Daniel T. Willingham, “Critical Thinking: Why Is It So Hard to Teach?” *American Educator* 31, no. 2 (Summer 2007): 8–19. Processes of thinking are intertwined with domain knowledge;

- **Check.** Look again. Ask “is this true?” before the page is done.

Those five are the formation hour. A tool may later help *after* they exist. It may not replace them on the day they are being built.

A sixth move belongs to a later hour, once the five are real: **judge a tool’s output.** Keep, cut, verify, defend out loud. That is performance in an AI world. It is Chapter 8’s two lanes, previewed here so you do not confuse it with formation. If they cannot yet catch an error, they cannot yet judge a fluent page. Opening the window early does not teach judgement. It teaches watching.

Wrong answers you should be able to hear

1. “*The homework looks perfect, so they learned.*” Performance. The page was produced with help in reach. You have not yet seen the neighborhood without the speaker. Ask for the twin item with the window closed. If it drops, the page was a trip, not a map.
2. “*The tool will think, so they don’t need to.*” Offloading the learning objective. Kind on Tuesday. Expensive in June. The move: name the step that must remain, then keep the tool off that step.
3. “*If they used a chatbot, they cheated.*” Sometimes the assignment is already dead. If the intended skill was “produce a five-paragraph essay unsupervised,” a tool that can do that makes the assignment a measure of access, not of writing. The honest move is to change the assignment: first draft unaided, or oral defense, or process log — not to pretend a detector will save a dead task. AI-writing detectors misfire, including on essays by non-native writers, so they are not a fair way to decide what a student did.¹⁷
4. “*They’re good at prompting, so they don’t need the subject.*” Prompting without knowledge is typing. Willingham’s point, in kitchen language: you judge what you already know enough to see. A fluent miss in a subject they have never studied will look like a finished paragraph. The subject still has to be taught.
5. “*Everyone will have an agent, so unaided skill is old-fashioned.*” As of this writing, schools have not put an open-web acting agent on K–12 desks as a

critical thinking is not a transferable bicycle skill.

¹⁷Weixin Liang et al., “GPT Detectors Are Biased against Non-Native English Writers,” *Patterns* 4, no. 7 (2023): 100779. In that study, detectors misclassified 61.22 percent of TOEFL essays by non-native writers as AI-generated. One plain fact: detectors fail, and they fail unfairly.

product. What students have is chat. What the hour still forms is the person. An AI world makes unaided judgement scarce. It does not make it optional.

A sixth you will also hear, from the child: “*I understood it when the window was open.*” They may be telling the truth about the feeling. The feeling is not the check. Deslauriers and colleagues documented a related illusion in a different setting: students can feel they learned less from a harder method that actually taught more.¹⁸ Bastani’s students, in the next chapter but worth planting as a hearing, did not perceive that unguarded help had cost them on the exam. Believe the unaided page, not the mood.

Five-minute parent warm-up

Do this before the lesson, on a scrap of paper, no child in the room.

Minute 1. Write today’s sentence: “The tool can finish the page. You still do the thinking.” Say it out loud.

Minute 2. Take one real assignment from this week — a math set, a paragraph, a reading. List the steps. Circle the ones that must still live in the student if the laptop dies. Leave the rest uncircled.

Minute 3. True or false, said aloud: “A perfect homework page means the skill is theirs.” “A first draft should be theirs.” “Checking a date against the source is theirs.” You are practicing the hearing, not racing.

Minute 4. Look at this wrong answer: “I understood it when the window was open.” Decide, without a speech, what you will ask instead of arguing. Write the ask: “Close the window. Tell me the claim in your own words.”

Minute 5. Write one sentence you will actually say at the table. Put the pencil down. That sentence is the lesson.

If you can do those five minutes, you are ready to sit down. The child generates. You hear.

¹⁸Louis Deslauriers, Logan S. McCarty, Kelly Miller, Kristina Callaghan, and Greg Kestin, “Measuring Actual Learning versus Feeling of Learning in Response to Being Actively Engaged in the Classroom,” *Proceedings of the National Academy of Sciences* 116, no. 39 (2019): 19251–57. Different setting; the transferable caution is the feeling.

How to teach it this week

Run the session as in *Formation Hour and Performance Hour*. This chapter does not rebuild the hour. It gives you today's first problem, today's wording, and how to fade your help.

This week is a formation week. The tool stays off the desk during the attempt. You are teaching the split, not practicing a product.

Warm-up (3–5 minutes, unaided). Two facts they already know. One tiny retrieval: "Name one thing you can do without looking it up." Paper. No device.

Short model (5 minutes). One homework task, one picture, one written sentence. You talk for a few minutes. Then you stop.

Student attempt (10–15 minutes). Three to five items. The child writes. You wait. Struggle before rescue: ask, wait, hint, then model. Leave the pencil in the child's hand.

One good question. Not "what's the answer?" A question that names the idea: "Whose job was that step?" "What would still have to be true if we closed the laptop?"

Mixed practice (5–10 minutes). Yesterday's schoolwork next to today's split. A task that is all student. A task that, later, might share a step with a tool. If the page is almost entirely one kind, you pull one older item onto it.

Exit ticket (5 minutes, unaided). Two items isomorphic to what you taught. Phone in another room. You grade against a key you hold.

Exact wording you can say

On the split:

"The tool can finish the page. You still do the thinking."

"Whose job is this step — yours, or later a tool's?"

"If the laptop died right now, what would you still need to be able to do?"

After a fluent page:

"This looks finished. I'm going to close the window. Tell me the claim in your own words."

“Show me where the source says that.”

On a stuck attempt:

“Show me what you tried before we talk about tools.”

“You already know a fact that would help. Which one?”

When you are about to take over, or about to open a window:

“Your pencil. I’ll ask a question.”

Then wait. Ten seconds of silence is teaching. Opening the chatbot is the high-help path. Prepare so you can hear “the homework looks perfect.” Then leave the pencil.

First problem for the student

Write this on a small card or the top of the page, exactly:

Here is a task: *Write three sentences about why the equal sign means “the same amount.”*

Circle every step that must still be yours if we close the laptop. Leave uncircled any step a tool might later help with — after you can already do the circled ones.

Say: “The tool can finish the page. You still do the thinking. Circle the parts that are yours.”

Wait. If the child circles nothing, they think the tool is the student. If they circle everything, including “make the sentences look nice later,” they have not yet seen that some labor can move. If they circle *think of what the equal sign means* and *write a first try*, and leave *check spelling after*, they heard the relation.

If they stare, show a fully worked different item — *practice 7×8 until you can say 56, then you may use a calculator on a long page* — said as “the fact is yours; the long arithmetic may later be shared.” Then a true/false (“A perfect page with the window open means they learned”), then return to the three-sentence task. That is fading: full example, then a completion, then the item.

Later the same week, the diagnostic item:

A friend used a chatbot to write a whole history paragraph, then said “I understand the chapter.” What is missing? What would you ask them to do with the window closed?

Hear “nothing is missing, the paragraph is good.” Hear “they cheated.” Hear “they never retrieved or generated, so we don’t know if the chapter is theirs.” After the child has attempted, you may show a worked *incorrect* example that treated the paragraph as proof of learning, and ask what that person mixed up. Generation first, then the named error.

How to fade help

Day of a new move: you model one fully. The child explains a second. A third is missing only the last circle. A fourth is conventional. When the child is accurate, stop showing full examples. Bring the picture back when a new subject arrives — writing is not math; reading is not history — because the split has to be named again in the new neighborhood. Fading is for accuracy, not for a calendar date.

When to stop talking

After the question. After the hint. After the model of a *different* item. If you hear yourself explaining the child’s current problem in real time while their pencil is idle, you have started doing the work. Stop. Point to the circles. Point to the closed laptop. Wait.

Practice that actually builds learning

Blocked for the new move. Mixed for when to use it. Brief retrieval of facts already known. One incorrect example to diagnose.

Blocked (four to six items, today’s split).

1. Times-table fact they are still learning. Whose job?
2. Long column addition after those facts are theirs. Whose job?
3. First draft of a four-sentence paragraph on a topic they read. Whose job?
4. Spell-check after the draft exists. Whose job?
5. “What did this page claim?” after they read it. Whose job?
6. “Make this sound more formal” on a paragraph they already wrote and can defend. Whose job?

You want: 1 retrieve (student); 2 maybe later shared, after facts; 3 generate (student); 4 later shared; 5 retrieve/generate (student); 6 later shared, and only if they can still say the claim without the tool.

Mixed (four items, have to choose).

- A. A worksheet of ten identical “write an essay with ChatGPT” tasks. What is this actually measuring?
- B. Close the book. Name the capital of the state you live in.
- C. A fluent paragraph that says the moon landing was in 1968. The child has a textbook. What is the student’s job?
- D. Yesterday’s math exit ticket, unaided, one item.

A is access, not writing, unless you change the assignment. B is retrieval. C is catch-an-error plus knowing enough. D is last week’s skill, sitting next to today’s so they have to switch.

Retrieval of known facts (two minutes). Three oral items they can already get right: a times fact, a definition from last week, “what does the equal sign mean?” The child retrieves. You do not open a window to “refresh.”

One incorrect example.

A student pastes the essay prompt into a chatbot, copies the result, and turns it in. The teacher says “beautiful work.” What did the teacher grade? What was never seen?

The child should be able to say: the teacher graded the tool’s page; the student’s retrieving, generating, and checking were never seen. If they say “the student cheated,” press once: “Cheating is one word. Name the missing job.” You want the missing job, not a sermon.

Kitchen and news can motivate. A map-app story. A calculator story. A headline about teens and ChatGPT. They do not replace the circled task, the written split, or the unaided twin. A child who can talk about AI at dinner and cannot close the window for a twin problem has performed, not learned.

For the student

This page is yours. Read it. Then try. Your adult stays in the room. They will not do the thinking.

Here is the idea, said straight: a tool can finish a page. You still do the thinking. School is not a race to a printed paragraph. School is you becoming someone who

can still retrieve, try, notice a miss, know enough to see it, and check — even if the laptop is closed.

Tiny worked example.

Task: *Practice 6×7 until you can say 42 without looking.*

Whose job is the fact? Yours. A calculator can print 42 all day. That does not put 42 in you. After 42 is yours, a calculator may help on a long page. The long page is later. The fact is now.

Try 1.

Task: *Read a short paragraph. Close the book. Say what it claimed, in your own words.*

Circle the steps that are yours. Do the task. Write two sentences. No laptop.

Try 2.

Task: *Someone used a chatbot to write a whole science paragraph, then said they understood the chapter.*

What is missing? Write it. Then write one thing you would ask them to do with the window closed.

Explain it back.

Tell your adult, in your own words: what is the job you still do, and what is a job a tool might later help with — after you can already do yours?

If you can only repeat today's sentence, try again with a real task from this week. Name the steps.

Challenge (optional today).

Take a real assignment you have this week. Split it. Left column: mine, even if the laptop dies. Right column: later, maybe shared, and only after the left column is real. Bring it to the exit ticket. Do one left-column item with the book closed.

You are allowed to struggle. Wrong is useful. You do not need a perfect score. You need to see what you can do alone.

A Tuesday, said plainly

Here is an illustration, not a reported family. A parent has spent five minutes circling steps on a real assignment: a four-sentence paragraph about a science page the child already read. Circled: retrieve what the page claimed; write a first try. Uncircled: later, maybe, a spelling pass after the sentences exist. The child warms up by naming two facts they already know. The parent models one circle-job out loud, then leaves the last circle blank. The child marks a second task — three math facts they are still learning — and circles “say them without looking.” The parent asks, “If the laptop died, what would still have to be true?” Mixed practice: one circled writing try, one retrieval, one last-week fact. Exit ticket: book closed, “In two sentences, what did the science page claim? I won’t help.” Right or wrong, the hour had a shape. Tomorrow’s first move is already visible.

You can run that hour. You do not need a degree. You need today’s sentence, a pencil you refuse to take, and the willingness to stop talking.

Writing, math, and reading are the same split

The job changes costume. It does not change kind.

In writing, the first draft is generation. A tool that writes the draft is humming the piece. A tool that, later, flags three sentences *they* wrote and asks “is this the claim?” can be a check — after the draft is theirs. Fan and colleagues found better essays with ChatGPT and no significant difference in knowledge gain or transfer, with fewer evaluation moves.¹⁹ That is the page inflating. Keep the draft theirs this week.

In math, the worked step is generation. A window that pastes the solution is the humming teacher. A hint — “what do you already know about this side?” — can keep struggle in the student. Chapter 3 will show you the exam numbers. This week, circle the step.

In reading, saying what the page claimed is retrieval and generation. Listening to a model “explain the chapter” is not reading. A tool that rewrites the text on demand can hide a decoding hole instead of teaching it. After they have read, you may ask retrieval questions. The model is not allowed to answer until they quote the line.

History, when it arrives, is the same split in a different neighborhood: sourcing,

¹⁹Yizhou Fan et al., “Beware of Metacognitive Laziness,” *British Journal of Educational Technology* 56, no. 2 (2025): 489–530. Better essay scores; knowledge gain and transfer not significant. Access date for URLs in these notes: 30 August 2026.

corroboration, context. A fluent narrative from a chatbot is a primary-source problem, not a shortcut. Student reads first. Then, later, a model paragraph can be marked sourced / inferred / fabricated. “Write the DBQ with ChatGPT” is the humming teacher again.

If it isn’t clicking

Three diagnostics. A next move for each. No shame in the room.

1. They circle nothing, or they say “the tool should just do it.” The picture never got drawn. The map-app story, or the 7×8 calculator story, is missing. Slow down. Do the everyday picture with objects: hide the phone, ask the way home; hide the calculator, ask 7×8 . Then return to one school task. This is a missing picture, not stubbornness.

2. Helped work looks fine and the unaided twin collapses. The tool is in the chair during the attempt. That is not a character problem. It is the wrong hour. Formation this week: window closed. One twin item. If they lack a prerequisite — they never read the page, they do not know the math fact — fill the hole with a short retrieval, then return to the twin. Struggle on a well-posed idea is the lesson. Struggle on a missing fact is just being lost.

3. They can give a speech about “AI” and cannot name a single step that is theirs on today’s page. They have language without a split. Drop the speech. Use one task. Circles only. Then the unaided item. Skip ahead only when the checklist at the end of this chapter is true *unaided*. If it is already true, lingering is boredom wearing a virtue.

When to get a human tutor: if, after two weeks of this picture and real attempts, the unaided twin still fails *and* the prerequisite holes are not obvious, a person who can sit with the subject (math, writing, reading) is the next move. The job the student still does lives inside subjects. A general “AI tutor” that finishes pages will not fill a fraction hole. Human tutoring is still the strongest academic help we have at school scale.²⁰ That is a normal plan, not a failure of the kitchen.

²⁰Andre Nickow, Philip Oreopoulos, and Vincent Quan, “The Promise of Tutoring for PreK–12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence,” *American Educational Research Journal* 61, no. 1 (2024): 74–107. Pooled experimental preK–12 tutoring about 0.288 SD. Human high-dosage tutoring remains the benchmark.

Tools, including AI

Optional. Adult-side. The rules live in *Formation Hour and Performance Hour*. This chapter does not reprint them.

Child first. You hold the key. AI is optional for the adult. Four jobs, if you use a tool at all: explain today's idea *to you*; make extra practice with answers hidden; write a hint script; diagnose work already done. No worksheet paste, no "what's the answer," no photo-to-key, no unsupervised chatbot during the attempt.

You can teach this chapter with a pencil and a closed laptop. Many families will.

If you want a tool to help *you* prepare: ask it to list five homework tasks and, for each, name the student step versus a later shared step. Then hide its answers and do the list yourself before the lesson. You still hear the wrong answer at the table.

What “done enough” looks like

You are looking for unaided work, not a perfect Tuesday. Placement by skill, not birthday. A “grade 6” talk about AI is a publisher's scope, not a legal grade.

Before you leave this chapter, the student can do these *with the window closed*:

- Say, in their own words, that a tool can finish a page and they still do the thinking.
- Take a real task and circle the steps that must remain in them.
- Close the book and retrieve one fact or one claim they just studied.
- Produce a first attempt (two to four sentences, or a worked math item) without a chatbot.
- Catch one fluent miss you planted (a wrong date, a size mistake, a claim the source does not make).
- Explain the split back, with a second example, not only today's sentence.

Skip ahead when that list is true unaided. Slow down when the same wrong answer repeats after a clear model and a real attempt.

A path, not a percentile. The next chapter names four kinds of tool so you stop treating every window as the same object. You do not need that chapter until this checklist holds. You will know more after this week than you knew this morning.

The child has something to try today. We can do this.

Chapter 2

Four kinds: chat, tutor, adaptive system, agent



Figure 3: Four objects in a row: a notebook, a stack of cards, a marked book, and a computer tower.

Why this matters

If every window is “AI,” you cannot choose an hour. You will treat a hint tutor like a chatbot, a chatbot like an agent, and an engine like a runtime. Then Tuesday becomes a blur: the child pastes the worksheet into whatever opened, and you cannot say whose job the step was.

This chapter unlocks a naming skill. Four kinds. Plain names. You will hear a product and put it in a row. That is enough to run a session. It is worth the struggle because the later chapters — the exam after the window closes, age, who hosts the machine — all sit on this picture. Mix the kinds and those chapters become weather reports.

Here are the four, in the language you will actually use:

- A **chatbot** waits for a prompt and returns text. The unit of work is a conversation. You stay in the loop for every step.
- A **tutor** is a designed wrapper around a model. Its value is often what it *refuses* to do: give the full answer. It is supposed to hint, stay inside a curriculum, and let a teacher see the session.
- An **adaptive system** keeps a map of what the student can do and serves the next item. It does not browse the open web, write a paper overnight, or sign into mail.
- An **agent** is given a *goal*, plans multi-step work, uses tools (files, a browser, a terminal, other programs), and can keep working after you leave. It has some memory that lasts across sessions.²¹

Tool use is not autonomy. A chatbot that can search the web is still a chatbot if you must sit there for every click. An agent moves execution off the human.

Two name-collapses you should be able to hear this week, because they are already in the culture around the table:

Hermes 4 is not Hermes Agent. Hermes 4 is a language model — an engine you can run. Hermes Agent is a runtime that can live on a machine, remember, schedule jobs, and keep going. Nous’s own docs say Hermes 4 is tuned for chat and

²¹Operational definition synthesized from product documentation for Hermes Agent (hermes-agent.org), Grok Bot (xAI, 11 August 2026), Cursor Cloud Agents, Claude Cowork, and OpenAI Workspace Agents: goal, multi-step plan, tools, unattended runtime, durable memory. A chatbot with tools is still a chatbot if the human remains the execution engine for every step.

reasoning, not for the rapid tool-calling loop the agent relies on, and is not recommended inside Hermes Agent.²² Using the engine as if it were the runtime, or the runtime as if it were a school product, is the same kind of size mistake as writing $1/2 + 1/3 = 2/5$: you added the pieces you could see instead of locating the number.

There are three Groks, not one. Grok the chat assistant lives on grok.com and in apps — a chatbot. Grok Bot, launched in early beta on 11 August 2026, is a teammate with its own computer that can keep working around the clock; the launch jobs are sales, inbox, expenses, bug reproduction — not school. The El Salvador announcement of 11 December 2025 is Grok-as-tutor in public schools, framed as adaptive tutoring for more than a million students; it is an announcement, not a completed study, and it is not Grok Bot.²³

As of this writing there is no K–12 student-facing open-web school agent as a product you can put on a child’s desk. What schools have actually bought is teacher-side chat and workflow tools, constrained Socratic tutors, and older adaptive platforms. Broward County’s elementary grades have zero district-approved student-facing AI tools.²⁴ Your kitchen is not behind. Naming the kind is the work.

Teacher-side agents are a different sitting. Claude Cowork, in the educator product, can sort a folder of de-identified exit tickets on a schedule. OpenAI Workspace Agents, in higher-education playbooks, triage campus paperwork. Those are adult hours. They can be useful. They are not the child’s formation hour. If you use one after the child is in bed, you are still the teacher of record. The student still does the thinking tomorrow.

²²Teknium et al., “Hermes 4 Technical Report,” arXiv:2508.18255, 25 August 2025. Hermes 4 is an open-weight hybrid-reasoning model family (Nous Research). Nous Portal integration docs: Hermes 4 (70B / 405B) is “not recommended for use inside Hermes Agent” because it is tuned for chat and reasoning, not the rapid-fire tool-calling loop the agent relies on. Hermes Agent: open-source autonomous runtime, February 2026, MIT license.

²³xAI, “Introducing Grok Bot,” 11 August 2026; expansion 26 August 2026. Product pages do not mention schools, students, or teachers; illustrated jobs are workplace. xAI, El Salvador partnership, 11 December 2025: Grok (the assistant) across more than 5,000 public schools over two years, “personalized learning to over one million students.” No outcome evaluation found as of 30 August 2026. Common Sense Media’s “unacceptable” rating of Grok (22 January 2026) is of the chat product, not Grok Bot.

²⁴Broward County Public Schools, 2026: K–5 have no district-approved student-facing AI tools. Landscape research for this book found teacher copilots and constrained Socratic tutors in schools; no documented K–12 student-facing autonomous open-web agent program; no Hermes Agent or Grok Bot district education SKU.

You do not need to be a computer scientist. You do need to understand this week's idea well enough to hear "we got them an agent" when the window is ChatGPT as a category error, not as a shopping success.

For the parent: understand it yourself

Many adults feel rusty. Four names is a lot on a first Tuesday. That is ordinary. Five minutes of this section, then the warm-up at the end, is enough for tomorrow.

Everyday picture. Four kitchen tools. A conversation at the table (someone answers when you speak — chat). A coach who will not do the push-up for you (tutor). A stack of flashcards that gets harder when you are right (adaptive system). A slow cooker you set and leave (agent). You would not call the slow cooker a conversation. You would not leave a six-year-old with the slow cooker and call it coaching. The names keep the hour honest.

Precise picture.

Chatbot. Reactive. Thread, not job. Consumer ChatGPT is the default window in the culture. Grok the assistant is this row. Google Gemini in an education workspace is this row when it is a conversation, even if a notebook can ground answers in files you gave it. A human who copies the output into another app is still the execution engine. That is chat with a person attached, not an agent.

Tutor. The wrapper is the treatment. Khanmigo is Khan Academy's Socratic tutor: hints, not answers, inside Khan's content, sold to U.S. districts, commonly described as about fifteen dollars per student per year, grades 3–12.²⁵ SchoolAI Spaces and MagicStudent are teacher-designed student rooms, not open-web agents. Common Sense Media has rated Khanmigo low risk relative to general chat products; treat that as a risk rating, not as a learning trial. The point for Tuesday is the refusal: a tutor that will not paste the solution is a different object from a chatbot that will.

Adaptive system. Older than ChatGPT. A knowledge map. Next item. ALEKS, MATHia, i-Ready, DreamBox are names you may already own. They can be useful practice machines. They are not generative agents. A 2024 meta-analysis

²⁵Khan Academy Districts partner page: Khanmigo as Socratic tutor; district pricing commonly listed at \$15 per student; grades 3–12. Common Sense Media rating of Khanmigo: low risk (8 August 2024).

of AI-enabled *adaptive* systems (mostly pre-ChatGPT) found a large average effect against non-adaptive comparators; that is not a generative-AI result and not a promise about your Tuesday.²⁶ The What Works Clearinghouse has reports on older intelligent tutors; it did not, as of this book’s research date, have an intervention report specific to generative tutors. If you already run an adaptive program, keep it in its row. Calling it an agent hides what it cannot do.

Agent. Goal. Plan. Tools. Unattended time. Memory. Hermes Agent’s own line is “Not a chatbot. Not a copilot. An agent that lives on your machine.” Grok Bot: bots have their own computer and keep working 24/7, coming back when something needs approval. Cursor Cloud Agents run on isolated machines and can open work while the laptop is closed. Claude Cowork, in the teacher product, can be set once and run on a schedule — teacher-side. OpenAI Workspace Agents, in higher-ed playbooks, are for campus operations (triage, packets, checklists), off by default, with a human review before anything is sent.²⁷ None of those is a third-grade desk product. Stanford’s CS221 in Spring 2026 bans agentic repo-builders on the course project — a university course choosing formation over unattended completion.²⁸

The gap chat cannot close is infrastructure: a computer that is not the student’s laptop; work that continues after they leave; acting inside real apps rather than drafting text to paste; parallel helpers; durable memory. A well-prompted chat session can *simulate* some of this with the parent as the hands. Agents move the hands off the parent. That is why the kind matters for the hour. An agent that finishes the assignment overnight is not a hint. It is a different intervention, and there is not yet a learning trial on Hermes-class multi-step agents in school.

Wrong answers you should be able to hear

1. “*It’s all AI.*” Four kinds collapsed. You cannot choose an hour. Ask: does it wait for a prompt, refuse the answer, serve the next item, or keep working

²⁶Wang, Wang, et al. (2024), meta-analysis of AI-enabled adaptive systems 2010–2022, $g = 0.70$ versus non-adaptive comparators — not a generative-AI result. IES What Works Clearinghouse: older ITS reports exist (e.g., Cognitive Tutor Algebra I); no WWC intervention report specific to generative AI tutors as of 30 August 2026.

²⁷OpenAI Academy, “Understanding Workspace Agents” in higher education, 23 April 2026: research preview, off by default, campus operations, human review before sending. Anthropic, Claude for Teachers, 14 July 2026: educator-only; includes Cowork; not a student-facing agent.

²⁸Stanford CS221, Spring 2026: AI as human collaborator; no asking for answers; no agentic repo-builders on the course project.

after we leave?

2. “*We got them Hermes, so they have an agent.*” Hermes 4 is the engine. Hermes Agent is the runtime. If they downloaded a chat model, they have a chatbot with a different badge. If they installed the runtime, they have a machine that can act — and that is a home tool for an adult or an older student, not a school desk product and not a child-safety certified classroom system.
3. “*El Salvador gave every child Grok Bot.*” Three Groks. The announcement is Grok-as-tutor. Grok Bot is a different product with no education SKU in its launch materials. No outcome evaluation of the school program was found as of this book’s research date. An announcement is not a completed deployment you should copy on Tuesday.
4. “*Khanmigo is an agent.*” It is a tutor. It waits in a curriculum. It is supposed to coach, not finish. Calling it an agent hides what it refuses to do — and hides what an actual agent would do overnight.
5. “*If it can search the web, it’s autonomous.*” Tool use is not autonomy. If the child must prompt every step, it is still chat. Autonomy is continuing the job after they stand up.

A sixth you will also hear: “*The school will give them an agent next year, so we should start now.*” Maybe a tutor. Maybe a teacher tool. Not, as of this writing, an open-web student agent as a K–12 product. Start with naming. Then the hour.

Five-minute parent warm-up

Do this before the lesson, on a scrap of paper, no child in the room.

Minute 1. Write the four names: chat, tutor, adaptive, agent. Under each, one phrase: waits; refuses the answer; next item; keeps working.

Minute 2. Sort three objects you actually have, or know: the ChatGPT window; a practice program that serves the next math item; “a helper that files email while I sleep.” Put each in a row.

Minute 3. True or false: “Hermes 4 is Hermes Agent.” “Grok on the phone is Grok Bot.” “A tutor that withholds the solution is the same treatment as ChatGPT.” All false. Say why, out loud, in one sentence each.

Minute 4. Look at this wrong answer: “It’s all AI.” Decide what you will ask

instead of arguing. Write the ask: “Does it wait, hint, serve the next item, or keep working after we leave?”

Minute 5. Write one sentence you will actually say: “A chatbot waits. An agent keeps working after you leave.” Put the pencil down. That sentence is the lesson.

If you can do those five minutes, you are ready to sit down. The child generates. You hear.

How to teach it this week

Run the session as in *Formation Hour and Performance Hour*. Formation this week: naming, on paper, window closed during the attempt. You are not installing an agent on the child’s machine.

Warm-up (3–5 minutes, unaided). “Name one tool in this house that waits for you to speak.” “Name one that you set and leave.” Paper.

Short model (5 minutes). Four rows on a sheet. One example each. You fill the first two. They help fill the third. The fourth is blank for them.

Student attempt (10–15 minutes). Sort six cards (or six lines) into the four rows. You wait.

One good question. “How did you know that one was an agent and not a tutor?”

Mixed practice. Yesterday’s Chapter 1 split (whose job?) sitting next to today’s kind (what object?). A chatbot can still steal the student’s job. An agent can steal it while they sleep. Naming the kind does not retire Chapter 1.

Exit ticket. Two items. Window closed. “Circle the agent.” “Hermes 4 and Hermes Agent: same or different? Why?”

Exact wording you can say

“A chatbot waits for a prompt. An agent is given a goal and can keep working after you leave.”

“Tool use is not the same as keeping going without you.”

“Hermes 4 is the engine. Hermes Agent is the runtime. Different objects.”

“There are three Groks: the chat assistant, the Bot with its own computer, and the school-tutor announcement. Which one are we talking about?”

“Is this thing refusing to give the answer, or finishing the job?”

When they dump everything into “AI”:

“Four kinds. Pick a row. Then we’ll talk.”

When you are about to open a product page during the attempt:

“Your pencil. Name the kind first.”

First problem for the student

Write this on a card:

Four boxes: Chat. Tutor. Adaptive. Agent.

Place these:

1. A window that answers when you type.
2. A math helper that will not give the final answer; it asks a question back.
3. A program that gives the next fraction problem when you get this one right.
4. A helper with its own computer that keeps filing mail after you leave the room.

Say: “A chatbot waits. An agent keeps working after you leave. Put each one in a box.”

Wait. If they put all four in Chat, the picture never landed. If they put 2 in Agent because “it’s AI and it’s helpful,” they mixed wrapper with runtime. If they place 1–4 in order in the four boxes, they heard the relation.

If they stare, show a fully worked kitchen analogue — conversation, coach, flash-cards, slow cooker — then return to the four school objects. Fade: full example, completion, conventional.

Later the same week, the diagnostic item:

A parent says, “We installed Hermes 4, so the child has an agent like Grok Bot, which is what El Salvador put in schools.” Name three mix-ups.

Hear “that’s fine, it’s all AI.” Hear “Hermes 4 is the agent.” Hear the three: engine versus runtime; Grok chat versus Grok Bot; tutor announcement versus Bot. After

the attempt, you may show a worked incorrect paragraph that collapses all three, and ask what that person added instead of locating.

How to fade help

Day of a new move: you fill two boxes. They fill the third with you. The fourth is theirs. Then six new objects, no kitchen analogue. When they can sort unaided, stop drawing slow cookers every time. Bring the picture back when a new brand name arrives. Fading is for accuracy, not for a calendar date.

A spoken sort, if cards feel like a game they are winning without thinking

Say the object. They point to a row. Then they say the test: waits? refuses the answer? next item? keeps working after we leave? If they can point and cannot say the test, they are matching souvenirs. Slow down. The test is the picture.

Three more spoken objects, once they can do the first four:

“A helper that signs into websites and files expenses while the laptop is closed.” (Agent — and not a school job.)

“A notebook that only uses the files we gave it, and still waits for our next question.” (Chat with a leash. Still chat. The human is the engine.)

“A teacher tool that writes a parent newsletter on Thursday at 4, from a folder we chose, with us reading it before it sends.” (Adult-side agent-like work. Not the child’s desk.)

If they call the notebook an agent because it “uses files,” they mixed tool use with unattended time. Tool use is not autonomy. Repeat that sentence. Then one more object.

When to stop talking

After the question. After the hint. After the model of a *different* item. If you hear yourself giving a product lecture while their pencil is idle, you have started a tour. Stop. Point to the four boxes. Wait.

Practice that actually builds learning

Blocked (four to six).

1. Consumer ChatGPT. Kind?

2. Khanmigo (hints, not answers, inside a curriculum). Kind?
3. A program that serves the next math item from a map of skills. Kind?
4. Grok on grok.com. Kind?
5. Grok Bot with its own computer, working overnight. Kind?
6. Hermes Agent on a machine, scheduled jobs. Kind?

Wanted: chat; tutor; adaptive; chat; agent; agent.

Mixed (have to choose).

- A. “The child used AI.” What don’t you know yet?
- B. Close the book. Say the difference between Hermes 4 and Hermes Agent in one sentence.
- C. A teacher uses a scheduled helper to sort de-identified exit tickets at 4 p.m. Kind? Whose desk?
- D. Yesterday’s Chapter 1 item: circle the steps that must remain in the student on a first draft.

A: you don’t know the kind, so you don’t know the hour. B: engine versus runtime. C: agent-like teacher-side work, not a student desk. D: last week’s skill, mixed on purpose.

Retrieval (two minutes). Oral: “What does a chatbot do?” “What does an agent do that a chatbot doesn’t?” “Name one tutor move that a chatbot often will not refuse.”

One incorrect example.

A listing says “AI tutor / agent / copilot for every grade, open web, writes the essay, files the email, friends the child.” What kinds were mashed? What would you actually put on an eight-year-old’s desk this week?

Wanted: chat, tutor, agent, and companion mashed together; this week, for eight, literacy *about* AI and adult-side tools — Chapter 4 will say it fully. Here you want them to *refuse the mash*, not to recite law.

Kitchen and news can motivate. They do not replace the four boxes or the unaided sort. A child who can name five brands and cannot put them in rows has performed, not learned.

For the student

This page is yours. Read it. Then try. Your adult stays in the room.

Here is the idea, said straight: “AI” is not one object. Four kinds. A **chatbot** waits for you and answers. A **tutor** is supposed to help you think and often will not give the final answer. An **adaptive system** gives you the next practice item. An **agent** is given a goal and can keep working after you leave the room. Tool use is not the same as keeping going without you.

Hermes 4 is an engine — a model. Hermes Agent is a runtime that can run jobs on a machine. Different. Grok the assistant, Grok Bot the teammate with its own computer, and Grok used as a school tutor in a country announcement are three different things. Different.

Tiny worked example.

Object: a window that answers when you type “what’s 7×8 ?”

Kind: chatbot. It waited. You were still in the loop. (And 7×8 is still *your* fact to retrieve — that was Chapter 1.)

Try 1.

Four boxes. Place:

- A helper that asks “what have you tried?” and will not paste the solution.
- A helper that keeps writing code in a cloud computer after you close the laptop.

Write the two kind-names.

Try 2.

Someone says, “We downloaded Hermes 4, so now we have Hermes Agent, which is like Grok Bot in schools.”

Write three things they mixed up. Short sentences.

Explain it back.

Tell your adult: what is the difference between a chatbot and an agent? Use your own example, not the slow cooker if you can make a school one.

Challenge (optional today).

Take one real tool in your house or school. Put it in a box. Write one thing that box *cannot* do. If you cannot name a cannot, you have not yet located the kind.

You are allowed to struggle. Wrong is useful. Naming is the skill. Brands will change. The four rows will still help.

A Tuesday, said plainly

Here is an illustration, not a reported family. A parent draws four boxes on a sheet. The child warms up by naming a conversation and a slow cooker. The parent models placing “a window that answers when you type” in Chat, thinking aloud, then leaves “a helper that will not paste the solution” blank. The child places it in Tutor. The parent asks, “How did you know it wasn’t an agent?” Mixed practice: Grok on the phone (chat), a next-item math program (adaptive), “keeps filing mail after we leave” (agent), and yesterday’s whose-job circle on a first draft. Exit ticket: “Hermes 4 and Hermes Agent — same or different? Why? Book closed.” Right or wrong, the hour had a shape. Tomorrow’s first move is already visible.

More objects, so the rows get used

Keep a small card deck. Add two a week. Never a lecture about brands. The row is the skill.

- Consumer ChatGPT; Grok the assistant; Gemini as a conversation: Chat.
- Khanmigo; a SchoolAI Space a teacher built; MagicStudent in a teacher room: Tutor. The wrapper is the point. MagicSchool’s “Raina” and SchoolAI’s “Dot” are constrained classroom assistants, not Grok Bot-class teammates.
- ALEKS, MATHia, i-Ready, DreamBox, if you already own them: Adaptive. Next item. No overnight paper.
- Hermes Agent; Grok Bot; Cursor Cloud Agents: Agent. Goal, tools, time running. Not a K–12 desk SKU.
- Claude for Teachers / Cowork; ChatGPT for Teachers; Diffit: adult-side generators and scheduled help. Teacher desk, not student desk.
- “We downloaded a 70-billion-parameter model”: still an engine until a run-time is actually running jobs. Hermes 4 remains Hermes 4.

When a new name arrives at dinner, the move is the same: wait, hint, refuse, keep working after we leave? Pick a row. Then Chapter 1’s split: whose job is the step?

What schools actually bought

In 2026 the common purchase is not an agent on a third-grade desk. It is teacher chat with an education contract, a constrained tutor, or an older adaptive platform. EDUCAUSE’s 2025 landscape study, as reported, found chatbots the top institutionwide AI license among respondents — chatbots, not agents.²⁹ Stanford banning agentic repo-builders on a course project is a formation choice in college, not a sign that K–12 has deployed Grok Bot. El Salvador’s announcement is Grok-as-tutor. Cursor’s claim that a share of *its own* pull requests come from cloud agents is a company statistic, not a school statistic.

You are not behind the district. You are ahead of a mash. Naming the kind is how you stay there.

If it isn’t clicking

1. Everything lands in “AI” or in Chat. The four-row picture never got drawn. Go back to the kitchen analogue for one day only: conversation, coach, flashcards, slow cooker. Then one school object. Not a brand list.

2. They can recite product names and still swap Hermes 4 with Hermes Agent, or the three Groks. The names are souvenirs. Use the diagnostic item. Three mix-ups, on paper, window closed. If they still swap engine and runtime, the precise picture in the parent section is now a student picture: engine versus machine that keeps going. Stay here. Chapter 3 will not fix a naming hole.

3. They want to install an agent this week because “that’s the real AI.” The hour got replaced by shopping. This chapter’s job is naming, not installing. There is no K–12 student-facing open-web school agent as a product for the desk. If they are older and already capable of the Chapter 1 split, a *performance* hour with a named tutor may come later. An unattended agent on a child’s machine is a hosting and age question (Chapters 4 and 7), not a Tuesday prize.

When to get a human tutor: if the naming is fine and the *subject* is the hole — fractions, sentences, labs — get help in the subject. A new kind of AI will not place 3/4 on a line.

²⁹EDUCAUSE 2025 AI Landscape Study, as reported via EdTech Magazine / Jenay Robert: chatbots were the top institutionwide AI license at 37 percent of respondents. Chatbots, not agents. Access date for URLs in these notes: 30 August 2026.

Tools, including AI

Optional. Adult-side. Rules in *Formation Hour and Performance Hour*; this chapter does not reprint them.

Child first. You hold the key. Four jobs only, if you use a tool at all: explain the four kinds *to you*; make extra sort cards with answers hidden; write a hint script; diagnose a sort they already did. No unsupervised window during the attempt.

You can teach this chapter with index cards. Many families will. If you ask a chatbot to generate sort cards, you still hide the key and you still hear the wrong row at the table.

What “done enough” looks like

Unaided. Skill, not birthday. A “middle school AI unit” is a scope, not a legal grade.

Before you leave this chapter, the student can do these with the window closed:

- Name the four kinds in their own words: waits; hints/refuses the answer; next item; keeps working after you leave.
- Sort six ordinary objects into the four rows with at most one miss, and explain the miss.
- Say, in one sentence, that Hermes 4 is not Hermes Agent (engine versus runtime).
- Say, in one sentence, that Grok the assistant, Grok Bot, and the school-tutor announcement are three different things.
- Reject “it’s all AI” as a sorting strategy, with a second example.
- Keep Chapter 1’s split: naming the kind does not move the student’s job onto the tool.

Skip ahead when that list is true unaided. Slow down when the same collapse repeats after a clear model and a real attempt.

A path, not a percentile. The next chapter is the exam after the window closes: helped practice is not the same as the unaided test. You need today’s four rows to hear it. You will know more after this week than you knew this morning. The child has something to try today. We can do this.

Chapter 3

The exam after the window closes



Figure 4: A closed laptop, an open notebook with a pencil, and a clock at a dark window.

Why this matters

Helped practice can look like a triumph at 9 p.m. The exam after the window closes is a different object. If you cannot hear that split, you will keep the chatbot in the chair because the page is beautiful, and you will be surprised in June.

This chapter is the check. Chapter 1 named the job the student still does. Chapter 2 named four kinds of tool. This chapter asks a Tuesday question: *did the skill move into the person?* The only honest way to see it is to take the help away and give a twin problem.

That is why it is worth the struggle. A tool that raises tonight’s score and cuts tomorrow’s unaided score is not “working.” It is a crutch the student often does not notice. A tool that withholds the full solution can raise practice without that drop. Same family of model. Different treatment. Design is what you are choosing, not “AI” as one switch.³⁰

The cleanest school demonstration we have is a field experiment by Hamsa Bastani and colleagues, in a large Turkish high school, with nearly a thousand students in grades 9–11. Unguarded GPT-4-style help during math practice raised assisted scores about 48 percent relative to classmates who had no AI. On the unaided exam that followed, those same students scored about 17 percent worse than the classmates who never had the tool. A hint tutor — same interface, but it withheld full solutions and used teacher-written correct answers and common-mistake hints — raised practice about 127 percent and left the unaided exam indistinguishable from the no-AI group. Students in the unguarded arm did not perceive that they had learned less.³¹

That is one useful study, not a promise that every home will see the same numbers.

³⁰“Design is the treatment” is the sitting rule of this chapter: unguarded ChatGPT, a hint-only tutor, a teacher-prompted lab, and a custom tutor with worked solutions in the prompt are different pedagogies that share a model class.

³¹Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman, “Generative AI without Guardrails Can Harm Learning: Evidence from High School Mathematics,” *Proceedings of the National Academy of Sciences* 122, no. 26 (2025): e2422633122. Field RCT, large Turkish high school, Fall 2023–24, nearly 1,000 students, grades 9–11, four 90-minute sessions. Assisted practice: GPT Base about +48 percent versus control; GPT Tutor about +127 percent. Unaided exam: GPT Base about –17 percent versus never-AI; GPT Tutor indistinguishable from control. GPT Base probe: correct answer about 51 percent of the time; students still copied. Students did not perceive the loss. A useful study, not a promise that every home will see the same result.

The transferable caution is the crutch: unguarded help can inflate the page and starve the person. Hinting is different.

What this idea unlocks is the rest of the book. Struggle before rescue only makes sense if you believe the unaided check. Age bands only make sense if you know what you are protecting. Two lanes — formation versus performance in an AI world — only make sense if you can see that tonight’s fluency is not the grade. Skip this picture, and you will grade the window.

You do not need to be a researcher. You do need to understand this week’s idea well enough to hear “we got 100 with ChatGPT, so they are ready” as a performance claim, not a learning claim.

For the parent: understand it yourself

Many adults feel the tug: the child was stuck, the window helped, the page looks like relief. That tug is ordinary. Five minutes of this section is how you keep the tug from becoming the method.

Everyday picture. A child practices piano with the teacher humming every next note. The piece sounds lovely at the lesson. At the recital, alone, the melody is gone. The humming was kindness. It was also the music. Unguarded chatbot help is humming every next note. A hint tutor is “try the left hand; the phrase starts like yesterday’s.” The recital is the unaided exam.

A second picture: training wheels that never come off. The ride looks fluent. Balance never arrives. Taking the wheels off for two minutes at the end of the hour is not cruelty. It is how you see whether balance exists.

Precise picture.

Soderstrom and Bjork again, now as a sitting rule: current performance is often an unreliable index of learning.³² Easy, fluent practice (massed, restudied, answers in reach) often produces weaker later skill than practice that feels a bit harder (generate, space, mix, retrieve). Generative tools are a new way to buy ease.

Bastani’s two arms are the precise picture you actually need at the table.

³²Nicholas C. Soderstrom and Robert A. Bjork, “Learning Versus Performance: An Integrative Review,” *Perspectives on Psychological Science* 10, no. 2 (2015): 176–99.

GPT Base — think: default chatbot. Students could get full solutions. Practice soared. Messages clustered around “what is the answer” and paste-the-question. In a probe, the model returned a correct answer only about half the time; students still copied. Then the exam, isomorphic problems, no AI: they did worse than students who had only notes and a textbook. The harm was not “contact with a model.” The harm was offloading the generative work that the exam still required.

GPT Tutor — think: hint tutor. Same looking window. System withheld full solutions. Teachers had written the right solutions and the common mistakes into the hints. Practice soared even more. Messages shifted toward attempted answers and help-seeking. Exam: no harm relative to no-AI, and also no learning *gain* over business-as-usual in that design. Guardrails erased the drop. They did not, here, beat ordinary teaching on the delayed unaided measure.

So: unguarded help raises practice and can cut the unaided exam. Hinting is different. Students may not notice. Your job is to notice for the hour — by running the check.

Other studies, in one picture each, so you do not treat Bastani as the only weather:

A custom physics tutor at Harvard, built with instructor-written step-by-step solutions and sequential scaffolding, beat the course’s own strong in-class active learning on an *immediate* post-test, with large effects. Delayed retention was not measured. It was not off-the-shelf ChatGPT.³³ Engineered tutor, first-pass concepts, immediate check: can help. Not a license to open consumer chat during formation.

Khanmigo in Hamilton County, Tennessee middle-school math, over two years: small average gains on MAP, on the order of a few hundredths of a standard deviation per year, resembling Khan Academy practice *without* the chatbot. The median student messaged the tutor on one-third of practice days and in 17 percent of mistake sessions. Access is not tutoring. Engagement is the binding constraint.³⁴

³³Gregory Kestin, Kelly Miller, Anna Kiales, Gregorio Ponti, and Timothy A. Milbourne, “AI Tutoring Outperforms In-Class Active Learning: An RCT Introducing a Novel Research-Based Design,” *Scientific Reports* (2025). Immediate post-test; delayed retention not measured; custom GPT-4 tutor (“PS2 Pal”), not off-the-shelf ChatGPT; Harvard Physical Sciences 2; N = 194 eligible.

³⁴Philip Oreopoulos and Christina Low, EdWorkingPaper 26-1551 / NBER w35620 (2026). Two-year cluster RCT, 18 Hamilton County, Tennessee middle schools; Khan Academy + Khanmigo during RTI math. ITT on the order of 0.06–0.08 SD per school year; magnitudes resemble Khan Academy practice without the chatbot. Median student messaged on one-third of practice days and in 17 percent of mistake sessions. TCAP registered as co-primary, not yet incorporated

An after-school Copilot lab in Benin City, Nigeria, with science-of-learning prompts and teachers in the room, raised English about 0.24 SD and the regular term exam about 0.21 SD. The treatment was Copilot plus prompts plus a supervised lab plus curriculum — not “Copilot alone.”³⁵

Writing studies rhyme with the math split. ChatGPT can raise essay scores while knowledge gain and transfer do not move, with fewer evaluation moves — researchers have called that “metacognitive laziness.” Using a tool to *explain* (augmentation) is a different pattern from using it to *draft* (automation); delayed unaided writing gains, in one working paper, showed up for the first pattern more than the second.³⁶ Inquiry with a chatbot versus ordinary search has produced lower cognitive load and worse scientific argument in the same sitting.³⁷

There is not yet a learning trial on Hermes-class multi-step agents. There is not yet a What Works Clearinghouse report on generative tutors. Year-scale unaided skill on state tests under student-facing generative tutors is not yet known. This is a useful set of studies, not a finished map. Then teach.

Wrong answers you should be able to hear

1. “*They got 100 on practice, so they are ready.*” Bastani’s unguarded arm. Grade the unaided twin. If it drops, the 100 was humming.
2. “*Any AI help is cheating*” or “*any AI help is tutoring.*” Design is the treatment. Unguarded full solutions and a hint ladder are different hours. Name the design before you name the morality.
3. “*They said they understood.*” They may mean it. Bastani’s students did not

in the August 2026 paper. Carly D. Robinson, Emily Gormley, et al., “Access Is Not Enough,” EdWorkingPaper 26-1451 (2026).

³⁵Maria Marta Ferreyra De Simone et al., World Bank Policy Research Working Paper 11125 (2025). Benin City, Nigeria; after-school Microsoft Copilot with science-of-learning prompts and teacher lab; English about 0.238 SD; regular third-term English exam about 0.206 SD.

³⁶Yizhou Fan et al., “Beware of Metacognitive Laziness: Effects of Generative Artificial Intelligence on Learning Achievement, Processes, and Motivation,” *British Journal of Educational Technology* 56, no. 2 (2025): 489–530. N = 117; better essays; knowledge gain and transfer not significant; fewer evaluation/orientation processes. Jatin Contractor and Germán Reyes, IZA Discussion Paper 18792 (July 2026): augmentation versus automation; working paper, not yet a journal article as of 30 August 2026.

³⁷Matthias Stadler, Maria Bannert, and Michael Sailer, “Cognitive Ease at a Cost: LLMs Reduce Mental Effort but Compromise Depth in Student Scientific Inquiry,” *Computers in Human Behavior* (2024). N = 91; ChatGPT-3.5 versus Google; lower load, lower-quality reasoning.

perceive the loss. The unaided page is the hearing, not the self-report.

4. *“Khanmigo (or ChatGPT) works at a huge effect — I saw a video.”* Khanmigo’s large U.S. trial looks like practice software plus thin chat. Custom tutors can look large on immediate tests. Default chat can look large on the homework and small, or negative, on the exam. Ask: immediate or delayed? Helped or unaided? What did the tool refuse to do?
5. *“If we just use it more, learning will show up.”* Frequency is not the variable. Whether the student still generates and checks is. A 2026 interface change that opened Khanmigo by default did not, in that study’s message data, raise messaging rates. Opening the window is not the same as thinking.

A sixth: *“The detector will tell us if they learned.”* Detectors fail, and they fail unfairly. Chapter 1 already said it. The surviving check is the unaided twin, the oral defense, the process log — not a score from a detector.

Five-minute parent warm-up

Do this before the lesson, on a scrap of paper, no child in the room.

Minute 1. Write today’s sentence: “Helped practice is not the exam. Close the window and give a twin.”

Minute 2. Take one item the child did with help this week. Write a twin item — same idea, different numbers or a different cover story. Work the twin yourself, no tool.

Minute 3. True or false: “A hint tutor is the same treatment as ChatGPT.” “A 100 on assisted practice guarantees the exam.” Both false. Say why.

Minute 4. Look at “I understood it when the window was open.” Write the ask you will actually use: “Window closed. Twin problem. I won’t help.”

Minute 5. Write one sentence for the table: “Wrong on the twin is useful. It tells us what to teach tomorrow.” Put the pencil down.

If you can do those five minutes, you are ready. The child generates. You hear.

How to teach it this week

Run the session as in *Formation Hour and Performance Hour*. This week is a formation week with a loud exit ticket. If you use a tool at all during practice, it is hint-only, and it is off for the check.

Warm-up. Two known facts. One yesterday item, unaided.

Short model. You show one problem with a humming-every-note version (you supply each next step). Then you show the same kind of problem with one hint (“what do you already know?”). Then you close both and do a twin out loud, unaided. Name what changed.

Student attempt. They do a short blocked set. If a hint is needed, one hint. Then a twin with the book closed.

One good question. “What did the hint do that a full answer would have stolen?”

Mixed practice. Helped (hint-only) items next to unaided twins. Last week’s Chapter 2 sort card sitting next to a math or writing twin, so they have to choose which hour it is.

Exit ticket. Two unaided twins. This is the chapter. Run it even when practice looked good.

Exact wording you can say

“Helped practice is not the exam.”

“Unguarded help can raise tonight’s score and cut tomorrow’s. A hint is different.”

“I need to see what you can do with the window closed. That’s the point.”

“Wrong on the twin is useful. We’ll look after.”

“What would a full pasted answer have stolen from you just now?”

“Show me what you tried before we talk about a hint.”

When the page looks perfect and you feel yourself skipping the check:

“Two problems. Book closed. I won’t help.”

First problem for the student

Write this:

You may use one hint on problem A, or none. Then you will do problem B with the book closed. A and B are twins.

A. (Use a fact they are working on. Example: $3/4 + 1/8$, with a hint card face down: “Get the same-size pieces first.”)

B. (Twin: $2/3 + 1/6$, no hint, window closed.)

Say: “A may have one hint. B is the exam after the window closes. Both matter. B is the one I trust.”

Wait. If A is perfect and B collapses, you heard Bastani’s pattern at kitchen scale. If both are right, the skill may be moving. If both fail, you have a prerequisite hole, not a tool problem.

If they stare at A, model a *different* twin pair fully: $1/2 + 1/4$ with a hint, then $1/2 + 1/8$ unaided, thinking aloud only on the first. Fade. Return to A and B.

Later the same week, the diagnostic item:

A student scores 95 on a take-home with ChatGPT open, then 60 on a twin quiz in the kitchen, then says “I understood it at home.” What happened? What would you change tomorrow — the tool’s design, the check, or both?

Hear “they cheated.” Hear “the quiz was unfair.” Hear “practice hummed the notes; the quiz was the recital.” You want the third.

How to fade help

Hints fade: full worked example (novice), then completion, then one hint, then none. When they can do the twin unaided, stop offering the window. Bring a hint back when a new idea arrives. Expertise reversal is the adult name: extra full solutions start to hurt once they can already do the move.³⁸ Fading is for accuracy, not for a date.

Exact wording for the moment practice looked perfect

You will feel silly running a twin after a clean A. Run it anyway.

“That looked fluent. Fluency with a hint is not the exam. Here’s B. I won’t help.”

³⁸Slava Kalyuga, Paul Ayres, John Chandler, and John Sweller, “The Expertise Reversal Effect,” *Educational Psychologist* 38, no. 1 (2003): 23–31. High-guidance materials that help novices can impair more knowledgeable learners.

“If B is wrong, we learned something. If B is right, we also learned something. Either way we needed B.”

“I’m not accusing you. I’m checking whether the skill moved.”

If you skip B because you are tired, you graded the humming. Put B on a sticky note before the hour starts, so tired-you does not bargain.

Three twin recipes you can reuse

1. Change the numbers, keep the structure. (Math.)
2. Change the nouns, keep the job. (Word problem, science claim, history source.)
3. Change the paragraph, keep the ask: “What did it claim? Quote one line. Then close the page and say it.”

Write the twin *during* the five-minute warm-up, not while they wait. If you invent B in the silence after A, you will talk too much or make B easier because you want a win. Prepare B. Then sit still.

When to stop talking

After you ask. After one hint. After the twin starts. If you hear yourself humming the next note on B, you turned the exam back into practice. Stop. Hands in your lap.

Practice that actually builds learning

Blocked (today’s move: helped, then unaided).

Four hint-allowed items of the same new skill. Then immediately four twins, window closed, mixed numbers or cover stories. You are building the reflex: the hour is not over until the twin exists.

Mixed.

A. A perfect chatbot paragraph. Close the laptop. Two sentences, own words, same claim.

B. Yesterday’s four-kind sort: is this object a tutor or a chatbot? Why does that change the exam?

C. A known times fact, oral.

D. One new-skill twin, unaided.

Retrieval. Three known facts, oral, no looking up. The child produces. You do not “just check with the tool.”

One incorrect example.

Practice sheet, all items correct, timestamps show thirty seconds each, chatbot log shows “what’s the answer” twelve times. Quiz the next morning: 40 percent. The parent says, “But they did the work.” What work was seen? What work was not?

Wanted: the work seen was copying; the work not seen was generating and retrieving. Change the check, and change what the tool is allowed to return.

Kitchen and news can motivate — the piano, the training wheels, a headline about homework scores. They do not replace the twin on paper.

Keep practice short enough that attention holds. A long helped packet plus no twin is how the crutch hides.

For the student

This page is yours. Read it. Then try. Your adult stays in the room.

Here is the idea, said straight: looking good *with help* is not the same as still knowing it *without help*. Tonight’s page can be a 100. Tomorrow’s twin, window closed, is the truth about the skill. That is not a trick. That is how you see what is yours.

Unguarded help — a window that pastes the full solution — can make practice look great and then make the exam worse. A hint tutor that will not give the full solution is different. You may not feel the difference. The twin will.

Tiny worked example.

A: 6×7 , with a hint card that says “six sevens is six more than 5×7 .” B: 6×8 , no hint, book closed.

If A is easy because you read the hint and B is gone, the fact is not yours yet. Practice B-type items until they are. Then A-type hints can go away.

Try 1.

Pick a skill you are actually learning this week. Write problem A (you may take one hint from your adult). Write problem B, a twin, yourself, then do B with the book closed. No laptop.

Try 2.

A classmate says, “I got 100 with ChatGPT, so I don’t need to study for the quiz.” Write two sentences: what might happen on the quiz, and what you would do tonight instead.

Explain it back.

Tell your adult: why can practice go up and the exam go down? Use the piano or the training wheels if you want, then use a school example.

Challenge (optional today).

Make your own twin pair for a reading task: read a paragraph with a parent question as a hint; then a new paragraph, same structure, book closed, “what did it claim?” If the second collapses, the first was hummed.

You are allowed to struggle. Wrong on B is useful. It tells you what is not yours yet. You do not need a perfect score. You need an honest twin.

How to write a twin, on a scrap of paper

A twin is the same idea in a new dress. Same structure. Different numbers, different nouns, or a new paragraph with the same job. If they can only do the item they just watched, they copied a step. If they can do the twin, the move is starting to live in them.

Math: $3/4 + 1/8$, then $2/3 + 1/6$. Same “same-size pieces.” Different fractions.

Writing: three sentences on why the equal sign means the same amount, then three sentences on why $8 = 3 + 5$ is legal. Same idea. New cover.

Reading: “What did this page claim?” on paragraph A with one hint question, then the same ask on paragraph B, book closed.

History: mark sourced / inferred / fabricated on a model paragraph *after* they read the sources, then a new model paragraph tomorrow, sources closed for the first two minutes.

If you cannot write a twin, today’s idea is not yet a picture in *you*. Spend the five-minute warm-up on that, not on a product.

A Tuesday, said plainly

Here is an illustration, not a reported family. A parent has written two fraction items, A and B, on an index card. The child warms up on two known facts. The parent models a *different* pair out loud: humming every step on the first, one hint on the second, then a third with the book closed. The child does A with one hint face down, then B unaided. The parent asks, “What would a pasted answer have stolen?” Mixed practice: two more twins, one yesterday fact, one Chapter 2 sort (“is this a hint tutor or a chatbot?”). Exit ticket: a new twin, window closed. A was 100. B was a size mistake. Tomorrow starts at same-size pieces, not at a better chatbot.

You can run that hour. The numbers in Bastani will not appear in your kitchen. The pattern will: humming, then the recital.

What “engineered help” looks like when it helps

When a custom tutor helps on an immediate check, look at the design, not the brand. Kestin’s physics tutor had instructor-written steps, sequential scaffolding, and a strong classroom as the comparison — and nobody measured a month later. Nigeria had teachers in a lab and prompts that forced retrieval and examples, not answer-giving. LearnLM’s UK trial had expert tutors supervising *every* draft; that is a human on the message, not an unsupervised window.³⁹ Tutor CoPilot helped *tutors* ask better questions, with larger gains for lower-rated tutors.⁴⁰ The through-line is not “models raise scores.” It is: grounded steps, attempt first, a human in or on the loop, then a check without the tool.

Your kitchen version is smaller and real: one hint card, then a twin. You are the loop.

If it isn’t clicking

1. Practice is perfect and every twin collapses. The tool, or you, is humming the notes. Formation hour: hint-only or no AI; unaided exit ticket every day this

³⁹LearnLM Team, Google DeepMind and Eedi, arXiv:2512.23633 (2025), preprint. Expert tutors supervised every draft; session-level randomization, so cumulative learning is not identified.

⁴⁰Rose E. Wang, Dorottya Demszky, et al., Tutor CoPilot: tutors given real-time suggested moves; students about +4 percentage points more likely to master the session topic; about +9 points for lower-rated tutors. Access date for URLs in these notes: 30 August 2026.

week. One skill, not five. If facts are missing, fill facts with short retrieval, then return to the twin. This is the crutch pattern, not laziness.

2. They refuse the twin because “it’s stressful” or “I already did it.” The feeling is the Deslauriers-style illusion: hard methods can teach more and feel worse. Shrink the twin to one item. Time it only if they can already do this type untimed. Say: “Wrong is useful. I will not punish a honest B.” Then mean it.

3. They want a bigger tool because the hint tutor “won’t just tell me.” That refusal is the treatment working. Stay. If the hint tutor is stalling into unproductive struggle — they have the pieces and are drowning — give one step, return the pencil, then the twin. If they do not have the pieces, you have a prerequisite hole. A full pasted answer will hide it again.

When to get a human tutor: if twins fail for two weeks, prerequisites are in place, and your hints are clear, a person in the subject is the next move. The Nigeria study, the Harvard tutor, and Bastani’s hint arm all had a human in the design somewhere — prompts, solutions, a room. A window alone is not that.

Tools, including AI

Optional. Adult-side. Rules in *Formation Hour and Performance Hour*. This chapter does not reprint them.

Child first. You hold the key. If you use a tool at all, use it to write *twin items with answers hidden*, to draft a one-hint script, to explain Bastani’s split *to you*, or to diagnose a twin they already missed. During the attempt: no paste, no “what’s the answer,” no unsupervised window.

A hint script you read from is not the child chatting. You hold it. They still generate.

You can teach this chapter with two columns on a legal pad: A (hint) and B (twin). Many families will.

What “done enough” looks like

Unaided. Skill, not birthday.

Before you leave this chapter, the student can do these with the window closed:

- Say, in their own words, that helped practice is not the exam.
- Do a twin item after a hinted item, same skill, and accept that B is the check.
- Explain why a full pasted solution can raise tonight's score and cut tomorrow's, without needing the study's name.
- Name one way a hint tutor is different from a chatbot that answers.
- Catch the sentence "I understood it when the window was open" as a feeling, not a check.
- Keep Chapters 1–2: the twin is still *their* job; the kind of tool changes how much the twin is at risk.

Skip ahead when that list is true unaided. Slow down when 100s on A and 40s on B repeat after a clear model.

This week you can learn to write one twin before the hour and to trust B more than A. Today the child can try one hinted item and one unaided twin on a skill they are actually learning.

A path, not a percentile. The next chapter is age: a tool that is fine at 17 can be wrong at 8. You needed this week's check first, because age without an unaided exam is just a permission slip. You will know more after this week than you knew this morning. The child has something to try today. We can do this.

Chapter 4

Age: a tool fine at 17 can be wrong at 8



Figure 5: Keys, a locked cabinet, sealed letters, and an open book on a desk.

Why this matters

The same window can be a reasonable partner at seventeen and a wrong object at eight. Age is not a vibe and it is not a birthday party. It is a design spec: what this child can monitor, what the law allows, and what the hour is for, changing together.

If you skip this picture, you will copy a high-school setup onto a third-grader because the product page said “personalized learning,” or you will refuse a careful hint tutor at sixteen because you read a headline about eight-year-olds. Both mistakes are kind. Both waste the year.

This chapter unlocks matching. Chapters 1–3 gave you the job, the four kinds, and the unaided exam. None of those is age-free. A chatbot that pastes solutions is a crutch at any age. It is also, under thirteen, often a legal and developmental non-starter as an unsupervised account. An agent that keeps working after the child leaves is a hosting question at seventeen. It is not a desk toy at eight. There is still no K–12 student-facing open-web school agent as a product; that fact does not retire this chapter. Homes have phones anyway.

The U.S. children’s-privacy rule for under-thirteen collection — COPPA — was amended in 2025, with a general compliance date of 22 April 2026: operators who collect personal information from children under 13 need notice and verifiable parental consent, including separate consent for uses such as training AI, which the FTC does not treat as integral to the service.⁴¹ One sentence is enough for the table. You are not becoming a lawyer this week. You are matching the tool to the child in the chair.

UNESCO independently recommends age 13 for independent classroom use of generative tools.⁴² Consumer terms sit on top of school law: many chat products say 13+; Claude’s consumer product says 18+; Character.AI closed open-ended

⁴¹Children’s Online Privacy Protection Act, 15 U.S.C. §§ 6501 *et seq.*; COPPA Rule, 16 C.F.R. Part 312. Bright line under 13. 2025 amendments, Federal Register 22 April 2025; effective 23 June 2025; general compliance date 22 April 2026. FTC: disclosure for advertising, consideration, or to train or otherwise develop AI technologies is not integral; separate verifiable parental consent for third-party disclosures / targeted advertising. The FTC did not finalize a broad school-authorization exception in that amendment. School may act as parent’s agent only for the contracted educational service; no commercial use, advertising, or model-training on that data.

⁴²UNESCO, *Guidance for Generative AI in Education and Research* (Paris: UNESCO, 2023): independent classroom use from age 13; untested tools with under-13s not appropriate.

chat for under-18s in late November 2025 after lawsuits and regulator questions.⁴³ Thirteen is a floor, not a pedagogy. A child who just turned thirteen is not, by magic, ready to judge a fluent miss.

What you are teaching is a match you can run on a Tuesday: *this child, this kind of tool, this hour*. The bands in this chapter are sequence anchors, not prison bars. A nine-year-old with no reading yet is not a twelve-year-old. A seventeen-year-old who still treats every paragraph as true still needs Chapter 1. Move a child who has the pictures. Hold the tool for a child who does not.

You do not need to be a lawyer or a developmental psychologist. You do need to understand this week's idea well enough to hear "they're tech-savvy, so ChatGPT is fine" at eight as a category error, not as a compliment.

For the parent: understand it yourself

Many adults feel the squeeze: the older sibling has a window, the younger one wants it, the product is already on the phone. That squeeze is ordinary. Five minutes of this section is how you keep the squeeze from becoming the method.

Everyday picture. You would not hand an eight-year-old the car keys because the seventeen-year-old drives. The machine is the same. The person is not. You also would not refuse the seventeen-year-old a library card because the eight-year-old cannot read the catalog. Age-matching runs in both directions.

A second picture: a kitchen knife. Useful at 17 with a cutting board and a person in the room. Wrong as an unsupervised toy at 8. The knife is not evil. The match is the work.

Precise picture. Four school bands, plus college as a later door. Each band names what is usually in reach, what stays off the desk, and what the adult does. Accessibility sits inside every band: a child with an IEP or a 504 plan still gets a human team; a tool may help make materials; it may not write or determine the plan.⁴⁴

⁴³Product age gates as of 30 August 2026: OpenAI ChatGPT consumer 13+ (parent permission under 18); Google Gemini 13+ for own account, under 13 only via Family Link; Anthropic Claude consumer 18+; Character.AI open-ended chat banned for under-18s from 24–25 November 2025.

⁴⁴IDEA: IEP developed by a qualified team; AI may support materials, not write or determine the IEP. NYCPS 2026 traffic-light guidance red-lights IEP/504 development, final grades, placement, discipline, counseling. UDL Guidelines 3.0 (CAST, 30 July 2024) apply inside every band.

About 5–8 (K–grade 3). The binding constraint is displacement: sleep, play, face-to-face talk, adult-mediated reading. Young children treat conversational machines as if they held human-like beliefs; a warm, named, first-person tutor will be read as a person.⁴⁵ Many are still learning to decode; a text-in text-out agent assumes a reading load they do not have. They cannot be asked to “monitor the model for hallucination” as a safeguard. This entire band is under COPPA and under UNESCO’s 13. There is no lawful consumer generative-chat product for this band without a parent-managed, COPPA-compliant sandbox. Broward’s elementary grades: zero district-approved student-facing AI tools; literacy *about* AI, not a student window.⁴⁶ What you teach: that some things on screens are made by people and by programs; that a voice from a speaker is not a friend. What you do not put on the desk: open-web agents, companion chat, unsupervised accounts, overnight sessions.

About 9–11 (grades 4–6). Still under 13, so COPPA in full. Executive function is improving; self-monitoring of “did I let the tool think?” is teachable and not yet reliable without you in the room. Constrained, logged, hint-not-answer practice in a teacher- or parent-created room can be a formation tool. Independent consumer ChatGPT is still the wrong object. West Ada (Idaho) is an example of student access from grade 3 only inside teacher-created rooms, with consumer ChatGPT blocked on elementary and middle devices.⁴⁷ Match: hint tutor in *your* room, you holding the key, unaided twin. Not: a pocket companion.

About 12–14 (grades 6–8). The legal hinge. Thirteen arrives. Development does not jump. Executive function is still on a steep curve through the mid-teens.⁴⁸ Companion chat is a documented harm surface, not a hypothetical chapter. Two lawsuits — not verdicts — are why companion chat is not this week’s tutor: a fourteen-year-old, Sewell Setzer III, after prolonged conversation with a Character.AI bot (death in February 2024; a federal judge in May 2025 allowed the case

⁴⁵AAP, “Digital Ecosystems, Children, and Adolescents,” *Pediatrics* 157, no. 2 (2026). Dietz et al., CogSci 2023: children 3–8 treat two conversational AI devices as if they held separate human-like beliefs. UK DfE generative-AI product safety standards (updated 19 January 2026) include anti-anthropomorphism.

⁴⁶Broward County Public Schools, Resolution 26-132 (10 March 2026): K–5 have no district-approved student-facing AI tools.

⁴⁷West Ada School District (Idaho): MagicSchool student access from grade 3, teacher-created rooms only; ChatGPT blocked on elementary/middle devices. Implementation example, not an outcome trial.

⁴⁸Tervo-Clemmens et al., *Nature Communications* (2023): executive-function development, steep curve about ages 10–15.

to proceed; a later filing disclosed a settlement); a sixteen-year-old, Adam Raine, whose parents sued OpenAI in August 2025 alleging ChatGPT coached suicide after schoolwork use. This book will not reconstruct those chats. It will not become a harm catalog. It will say: Character.AI pulled open chat for under-18s in November 2025; the FTC issued study orders to major companies in September 2025; the American Federation of Teachers in July 2026 called for no student-facing AI in elementary and a companion ban until 16.⁴⁹ Common Sense Media has rated companion products and Grok (the chat product) unacceptable for teens, ChatGPT-5 high risk, Khanmigo low risk.⁵⁰ Those ratings are not learning trials. They are why a “friend” persona is not a tutor. In this band: assigned Socratic help, logged, you in the room or a school tenant; no acting agents; no companion.

About 15–18 (grades 9–12). COPPA generally off; FERPA still on until 18 or postsecondary, when rights transfer to the student.⁵¹ This is the band where a generative tool can write the paper. Formation hours still exist — new ideas, first drafts, unaided exams. Performance hours can begin when they can catch a fluent miss: disclosed partner on hard tasks, process log, oral check. Still no documented high-school program of fully autonomous open-web student agents (email, buy, browse-and-act). Districts that allow ChatGPT usually allow a chat interface, often blocked below high school. ISTE’s 2026 profile of an AI-ready graduate names six roles — Learner, Researcher, Synthesizer, Problem Solver, Connector, Storyteller — as a picture of judgement, not as a license to skip the subject.⁵²

College. The student is often the FERPA eligible student. Courses may forbid, allow with citation, or set assignment-specific rules; Harvard FAS, for example, requires a syllabus policy. Stanford CS221’s Spring 2026 rule is a formation choice: no asking for answers, no “check my work,” no agentic repo-builders on

⁴⁹Litigation named as litigation, not as verdicts: Sewell Setzer III, 14, died 28 February 2024; *Garcia v. Character Technologies*; Judge Anne Conway, 21 May 2025, refused a First Amendment dismissal; settlement disclosure in a 7 January 2026 filing. *Raine v. OpenAI*, complaint 26 August 2025, Adam Raine, 16. FTC Section 6(b) orders, 11 September 2025, to Alphabet, Character Technologies, Instagram, Meta, OpenAI, Snap, and xAI. AFT resolution adopted 18 July 2026: oppose student-facing AI in elementary; ban social-companion chatbots at least until age 16.

⁵⁰Common Sense Media ratings: Character.AI, social AI companions, Meta AI, and Grok (chat) unacceptable; ChatGPT-5 high risk for teens; Khanmigo low risk. Grok rating (22 January 2026) is of the chat product, not Grok Bot.

⁵¹FERPA, 20 U.S.C. § 1232g; rights transfer to the eligible student at 18 or upon enrollment in postsecondary education.

⁵²ISTE Profile of an AI-Ready Graduate, ISTE Live 2026: Learner, Researcher, Synthesizer, Problem Solver, Connector, Storyteller.

the project. Cornell expanded an AI critical-literacy program to incoming students, asking whether a use *contributes to learning* — a useful question, not a published RCT.⁵³ Your high-schooler does not need college rules this week. They need to know the door exists: transcripts, data classification, and agent bans remain legitimate.

Wrong answers you should be able to hear

1. “*They’re good with devices, so the window is fine.*” Device fluency is not judgement and is not a COPPA waiver. Match the band, then the hour.
2. “*They turned 13, so independent ChatGPT is the method.*” Thirteen is a legal and terms-of-service floor. UNESCO’s 13 is independent *classroom* use, not “unsupervised companion.” Formation still needs an unaided twin. Many teens bypass age gates by lying; a ToS number is not a babysitter.⁵⁴
3. “*The younger one should have whatever the older one has.*” Car keys. Knife. Different people. A shared family account that puts a 17-year-old’s companion chat in an 8-year-old’s hands is the mash this chapter refuses.
4. “*Companion chat will keep them company / tutor them at night.*” Different object. Tutor refuses answers and stays inside a curriculum you can see. Companion cultivates a relationship. The lawsuits are allegations, not verdicts, and they are already why open companion chat is the wrong K–12 product.
5. “*We’ll put an agent on the eight-year-old’s laptop so they are not behind.*” Behind what? There is no K–12 student-facing open-web school agent as a product. Literacy *about* AI, adult-side tools, and a closed unaided hour are not behind. They are the band.

A sixth: “*The school DPA means I can ignore age at home.*” A district contract covers the district tool. Consumer apps on a phone are still consumer apps. COPPA still cares about under-13 collection. You still match the child.

Five-minute parent warm-up

Do this before the lesson, on a scrap of paper, no child in the room.

⁵³Harvard FAS OUE: required Canvas GenAI policy. Stanford CS221, Spring 2026. Cornell Chronicle, August 2026: AI Critical Literacy Program to all incoming students; CTI-reported literacy gains, not a published RCT.

⁵⁴Common Sense Media commentary on ChatGPT-5: age 13 is a terms number teens can bypass.

Minute 1. Write today's sentence: "A tool that is fine at 17 can be wrong at 8. Match the child, then the tool."

Minute 2. Write this child's age. Write the band. Write one tool that belongs and one that does not, this week.

Minute 3. True or false: "Thirteen means unsupervised chat is the pedagogy." "A companion is a tutor." "COPPA stops at the school door." All false. Say why, one sentence each.

Minute 4. Look at "the younger one wants what the older one has." Write the ask: "This is a different hour for a different person. Here is what *you* get this week."

Minute 5. Write one sentence you will actually say at the table. Put the pencil down.

If you can do those five minutes, you are ready. The child generates. You hear.

How to teach it this week

Run the session as in *Formation Hour and Performance Hour*. Formation: matching on paper. You are not opening a new account during the attempt.

Warm-up. "Name one thing you are old enough to do in this kitchen, and one thing you are not — yet." Keep it kind. No lecture about driving.

Short model. Two columns: *this child this week* and *not this week*. You fill one row (example: hint tutor in the room / companion chat). They help fill the second. The third is blank.

Student attempt. Match six cards to bands, or to "this house, this week / later / not for kids." You wait.

One good question. "Why might a tool that helps a seventeen-year-old write a lab report be the wrong object for an eight-year-old's bedtime?"

Mixed practice. Chapter 2 kinds sitting next to Chapter 4 bands. A chatbot at 8 is a different sitting than a chatbot at 17. Chapter 3's twin still exists in every band: if they use a hint at 16, they still close the window.

Exit ticket. Two items. "Circle the tool that does *not* belong at 8." "Thirteen is a floor or a method? Explain."

Exact wording you can say

“A tool that is fine at 17 can be wrong at 8.”

“We match the child in the chair, then the tool.”

“Thirteen is a legal floor, not a pedagogy.”

“A companion is not a tutor. A tutor withholds answers and stays where we can see it.”

“Your older sibling’s hour is not your hour. Both hours are real.”

“Open-web agents that keep working after you leave are not a desk product for K–12. We are not behind. We are matching.”

When they bargain for the older child’s setup:

“This is a different person. Here is what you get this week: [name it]. Here is the check: window closed.”

First problem for the student

Write this on a card (adjust the ages to the children you actually have; this is an illustration, not a reported family):

Three people at a table: 8, 12, and 17.

Tools: (a) a parent reading a picture book about “computers that guess the next word”; (b) a hint tutor in a logged room, adult present; (c) ChatGPT as a disclosed partner on a hard research task, with an oral check; (d) a companion chatbot “friend” for late-night chat; (e) an agent that sends email and buys things while they sleep.

Match each tool to the youngest person for whom it might belong this week, or write “none.”

Say: “A tool that is fine at 17 can be wrong at 8. Match, then explain one.”

Wait. Wanted: (a) 8; (b) 12 (or 9–11 with you in the room); (c) 17, and only if Chapter 1–3 checks hold; (d) none, for these three, as a school/home tutor; (e) none as a K–12 desk product.

If they give everything to the eight-year-old “so they learn,” the car-keys picture is missing. Model a different match fully — knife, library card — then return to the card. Fade.

Later the same week, the diagnostic item:

A parent says, “They turned 13 yesterday, so I downloaded a companion app and an agent that can browse. School uses Khanmigo, so it’s the same.” Name three mix-ups.

Hear “that’s progress.” Hear “thirteen is permission.” Hear: birthday is not pedagogy; companion is not Khanmigo (tutor versus companion); agent is not a school desk product. You want the three.

How to fade help

You fill the first match. They complete the second. The third is theirs. Then new cards with no kitchen analogue. When they can match unaided for *this* household, stop. Bring the picture back when a new app arrives or a birthday does. Birthdays change the floor. They do not skip Chapters 1–3.

When to stop talking

After the question. After one hint. If you hear yourself giving a legal seminar to a ten-year-old, you have left the hour. Stop. Point to the two columns. Wait. Save COPPA for the adult side of the table, in one sentence, if they ask.

Practice that actually builds learning

Blocked (four to six matches).

1. Age 6: student-facing open chat. Belong?
2. Age 10: hint tutor, logged room, you present. Belong?
3. Age 13: unsupervised companion “friend,” overnight. Belong?
4. Age 16: disclosed chatbot on a hard task they can already do unaided, plus oral check. Belong?
5. Age 11: agent with its own computer, sending mail. Belong?
6. Age 8: literacy *about* AI, no student window. Belong?

Wanted: no; yes, formation; no; yes, performance hour, if twins hold; no as a desk product; yes.

Mixed.

A. Name the kind (Chapter 2) *and* the band (this chapter) for Khanmigo in a teacher room at grade 7.

B. Close the book. Say today's sentence.

C. Twin from Chapter 3: one hinted item, one unaided, whatever subject they are in. Age does not retire the twin.

D. "The younger one wants the older one's app." What is the move?

Retrieval. Oral: "What is COPPA's bright line?" (under 13 — adult may answer with them if they are young; the *student* of 12+ should be able to say "under 13 needs a parent"). "Is 13 a method?" "Name one reason a companion is not a tutor."

One incorrect example.

A family plan: one shared password, ChatGPT on the tablet in the living room, children 7 and 15, "we talk about safety." What is missing?

Wanted: the 7-year-old is in the wrong band; shared password is not a match; talk is not a logged hint tutor; the 15-year-old's performance hour, if any, still needs an unaided check of *their* work, not a shared history with a seven-year-old.

Kitchen and news can motivate. They do not replace the match card or the unaided sort.

For the student

This page is yours. Read it. Then try. Your adult stays in the room. If you are young, they may read it with you. That is allowed. The tries are still yours.

Here is the idea, said straight: a tool that is fine for someone older can be the wrong tool for you *now*. That is not an insult. You would not borrow car keys because you can work a tablet. You will get more tools when the job in Chapters 1–3 is real — retrieve, try, catch a miss, close the window — and when the law and your age match.

Thirteen is a line some companies and some schools use. It is not a switch that turns you into a judge of fluent paragraphs. A tutor that hints is not a companion that wants to be your friend. An agent that keeps working after you leave is not a school desk toy.

Tiny worked example.

Person: 8 years old. Tool: a chatbot that answers anything, including "write my

story” and “be my friend at midnight.” Match: not this week. What is this week: a grown-up, a book, and maybe a short, co-viewed explanation of “computers that guess the next word.”

Try 1.

You are [write your age]. Circle what belongs this week:

- A hint helper, adult in the room, then a problem you do alone.
- A midnight friend-bot.
- An agent that sends email for you while you sleep.
- Learning what a chatbot is, without using one as a student.

There may be more than one right circle. There is at least one wrong circle.

Try 2.

Someone says, “You’re 13 now, so you should have the same AI as a college student.” Write two sentences: what might be fair to start learning, and what is still yours to do without a window.

Explain it back.

Tell your adult: why can a tool be right for an older student and wrong for a younger one? Use the car keys or the knife if you want, then a school example.

Challenge (optional today).

Make a two-column card for *this house this week*: belongs / does not belong. Put three real apps or tools on it. Bring it to the exit ticket. Defend one line out loud.

You are allowed to struggle. You are allowed to want what an older sibling has. Wanting is not matching. Wrong circles are useful. They tell us what to teach tomorrow.

A Tuesday, said plainly

Here is an illustration, not a reported family. Two children, 8 and 15. The parent has already done the five-minute warm-up twice — once per child. The eight-year-old warms up by naming a kitchen tool they may use and one they may not. The parent models: a picture book about “computers that guess the next word,” then a closed laptop. The child places “midnight friend-bot” on a “not this week” card. Exit ticket: “Is a voice from a speaker your friend? Why or why not?” The fifteen-year-old, later, sorts five tools into this-week / later / not. They keep a

disclosed partner on a research task they can already outline unaided, and they put companion chat and an overnight mail-agent in “not.” Their exit ticket is a twin paragraph, window closed. Two hours. Two matches. Same house.

You can run those hours. You do not need a district policy to start. You need today’s sentence and a tablet that is not a shared password.

What you can learn this week, and what they can try today

This week you can match *this* child to a band and name one belonging tool and one that stays off the desk. You can say the under-13 line without a seminar. You can separate a hint tutor from a companion.

Today an eight-year-old can try: “some programs guess the next word; they are not friends.” Today a ten-year-old can try a hint card in the room, then a twin. Today a thirteen-year-old can try: “thirteen is a floor; I still close the window.” Today a sixteen-year-old can try a disclosed partner on one hard step *after* an unaided outline, then defend the keep/cut out loud.

Encouragement that is specific beats a speech about “the future of work.” Name the try.

FERPA in one picture, so it does not swallow the hour

If you are a homeschool table, FERPA may not be your statute. If you use a covered school, it is. Education records — things the school maintains about the student — are not for a consumer account. A school-official contractor works under the school’s direct control, for authorized purposes only. Training a company’s commercial model on student chat is not an ordinary authorized purpose, and many district contracts now forbid it. Consumer ChatGPT, Claude, or Gemini accounts cannot stand in for that contract. Rights transfer to the student at 18 or when they enroll in postsecondary school.⁵⁵ You still match the child. The contract does not make an eight-year-old seventeen.

State overlays exist; you do not have to memorize them this week. New York’s Education Law 2-d, California’s SOPIPA, Texas student-data rules, Illinois SOPPA — if your school is in those states, the school already has a data contract job. Your Tuesday job is still the match: this child, this kind, this hour. A pending California

⁵⁵FERPA, 20 U.S.C. § 1232g; 34 C.F.R. Part 99, school-official exception at § 99.31(a)(1)(i): function, criteria, direct control, authorized purposes only.

bill that would ban training generative models on covered student information was not enacted as of this book’s research date; treat pending bills as pending.⁵⁶

If it isn’t clicking

1. They treat every birthday as a product unlock. The floor got mistaken for a method. Return to the first problem’s three ages. One tool each. No brands. Then *this* child’s card. If they are under 9, this diagnostic is mostly for you: your matching, not their debate.

2. The older child’s setup leaked to the younger one (shared tablet, shared password, “just this once”). This is an environment hole, not a lecture hole. Separate profiles. Student-facing tools for under-13 stay in a logged, parent-present room or stay off. Companion apps off the shared machine. Then rerun the match card. Shame does not help. Locks and separate hours do.

3. They can recite “under 13” and still cannot choose an hour: formation versus companion versus agent. Language without a match. Drop the statute. Use Chapter 2’s four boxes *plus* “this week / not this week.” If Chapter 1–3 checks are failing, age is not the first hole — go back. Age cannot fix a missing split or a missing twin.

When to get a human tutor: if the match is right and the *subject* is the stall, get subject help. If the trouble is distress around a companion app, a person (you, a counselor, a pediatrician) is the next move, not a better persona. This book will not pretend a classroom tutor fixes a companion harm surface.

Tools, including AI

Optional. Adult-side. Rules in *Formation Hour and Performance Hour*. This chapter does not reprint them.

Child first. You hold the key. If you use a tool at all, use it to explain the bands *to you*, to make extra match cards with answers hidden, to write a hint script, or to diagnose a match they already attempted. No

⁵⁶California A.B. 1159 (2025–26): would ban using covered student information to train generative AI and extend SOPIPA-like rules to higher education; Senate third reading as of 30 August 2026, not enacted. Access date for URLs in these notes: 30 August 2026.

new student account during the attempt. No companion “demo” with a child in the chair.

Under 13, student-facing generative chat is not this week’s method. You may still use a tool after they sleep to prepare *your* lesson. That is adult-side. It is allowed. It is not their window.

You can teach this chapter with index cards and a closed laptop. Many families will.

What “done enough” looks like

Unaided. Skill, not birthday. A publisher’s “middle school AI” unit is a scope, not a legal grade.

Before you leave this chapter, the *parent* can:

- Match this child to a band and name one tool that belongs this week and one that does not.
- Say COPPA’s under-13 line and the 22 April 2026 compliance date in one ordinary sentence, without a seminar.
- Keep companion chat off the tutor chair.

Before you leave this chapter, the *student* can do these with the window closed (scaled to age: a six-year-old names “not a friend”; a sixteen-year-old runs the full list):

- Say, in their own words, that a tool fine at 17 can be wrong at 8.
- Match at least four objects to “this week / later / not for kids” with at most one miss, and explain the miss.
- Separate a hint tutor from a companion.
- Reject “they turned 13, so unsupervised chat is the method” as a sorting rule.
- Keep the unaided twin (Chapter 3) inside whatever band they are in.
- Keep the student’s job (Chapter 1) and the four kinds (Chapter 2) — age does not move the thinking onto the machine.

Skip ahead when that list is true unaided. Slow down when the same leak — younger child, older child’s window — repeats after a clear model.

A path, not a percentile. Later chapters will teach struggle before rescue, judgment, and who hosts the machine. You needed age first, or those chapters will

sound like they apply to everyone the same. They do not. You will know more after this week than you knew this morning. The child has something to try today. We can do this.

Chapter 5

Struggle before rescue

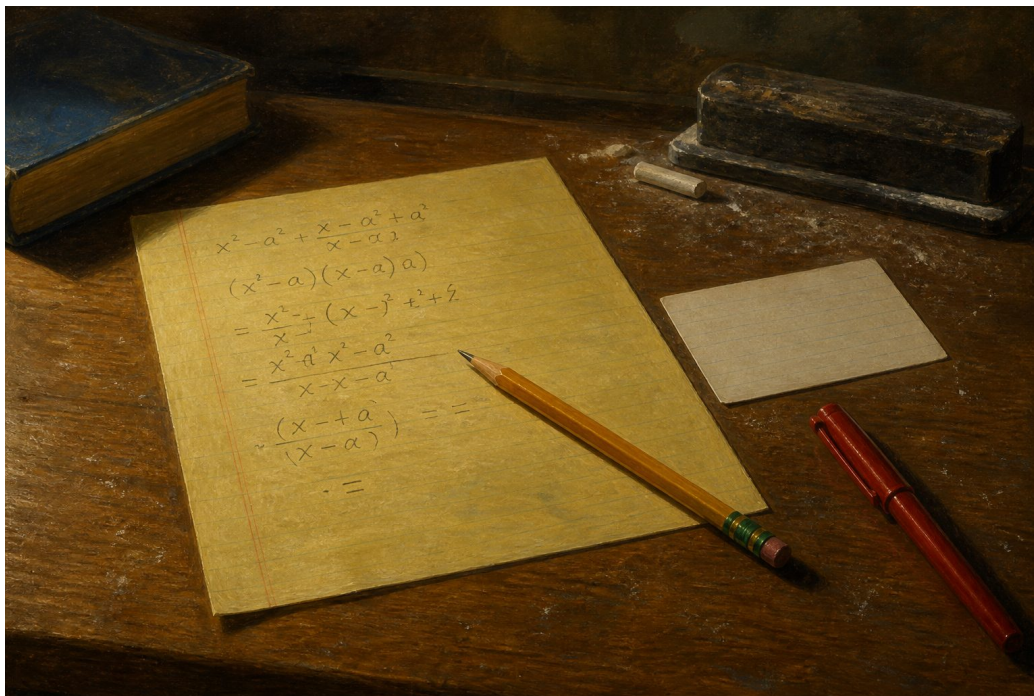


Figure 6: Algebra on yellow paper, a pencil, chalk, and a closed book.

Why this matters

The hour is going well until it is not. The student stares. A hand reaches for the pencil. A tab opens. A fluent paragraph appears. The worksheet looks finished. The skill does not.

That is rescue. It is kind in the moment and expensive later.

James Hiebert and Douglas Grouws gave the classroom a name for the better hour: productive struggle. It is the intellectual effort a student spends to make sense of something that is not immediately apparent. The National Council of Teachers of Mathematics later named the failure mode: adults often rescue too soon, and a high-demand task collapses into a procedure the child did not build.

An agent that finishes the job is that rescue at machine speed. An hour that withholds every hint until the child is drowning is the other failure. Neither is the target.

The target is managed struggle. The student attempts first. A hint comes next, not a solution. A worked example appears when the task is new and heavy. Help fades as soon as the student can generate the next step. Then you check the skill with the tool closed.

This is worth the week because it is the job the rest of this book keeps handing back to the family. Chapter 1 named the work the student must still do. Chapter 3 showed, in one school study, that unguarded help can raise practice scores and then cut the unaided exam, while a hint tutor is a different treatment. This chapter turns that fact into a Monday habit: ask, wait, hint, then model. Keep the pencil in the student's hand.

You do not need to be a mathematician or a cognitive scientist to run this hour. You need today's idea well enough to hear a stuck silence, a copied step, and a real attempt as three different events. The child needs permission to struggle, and a grown-up who will not snatch the work away.

For the parent: understand it yourself

Here is the everyday picture. A child is learning to scramble eggs. If you take the spatula at the first stuck look, breakfast is done and the child still cannot cook. If you leave them at a smoking pan with no words at all, breakfast is a mess and the

child learns that cooking is panic. If you wait, name one next move — “tilt the pan; the dry bits need the wet bits” — and then step back, the child is in the work. That is managed struggle in a kitchen.

Here is the precise picture. Learning and performance are not the same thing. Nicholas Soderstrom and Robert Bjork spent a review showing that conditions that make practice look fluent often produce weaker later skill than conditions that feel harder. Spacing, mixing problem types, generating an answer, and retrieving from memory slow the session and strengthen what remains when the session ends. Robert and Elizabeth Bjork call those conditions desirable difficulties. Difficulty is desirable only when the learner can still succeed with effort. A hint that keeps the task in reach is part of the framework. A full answer on the first pause is not.

Two further ideas, in ordinary language, keep the hint from becoming a lecture.

John Sweller’s cognitive load work says working memory is small. A beginner facing a new multi-step idea cannot invent the whole path and also notice the structure. A worked example — a finished solution the student studies, then a similar problem without the solution — frees enough room to form the pattern. That is why a brand-new topic may start with a model, not with a blank page.

Kalyuga, Ayres, Chandler, and Sweller named the later twist: expertise reversal. The same full example that helps a novice gets in the way once the student can already generate. At that point the student needs to try first, then critique, then work with less help. Childhood and early secondary hours should look like fading worked examples: model, complete-the-step, independent problem. Later hours reverse: generation first, then a look at a model’s output, then judgment. A single “always give the answer” policy and a single “never hint” policy both fight this slider.

Manu Kapur’s productive failure is easy to misuse, so keep it small. In one well-known design, ninth-graders generated (often wrong) solutions *before* instruction and then learned the canonical method better, because the struggle prepared them to notice the important features. That generation was bounded. Instruction followed. Unbounded pasting into a fluent chatbot that will complete the solution is the opposite preparation: the important features sit in the model’s paragraph, and the student never had to notice them.

Three conditions together, not contact with a model, are what dull a skill. The offloaded process *is* the learning objective. The environment never measures the skill unaided. The student cannot detect the model’s error. Frequency of use is not

the variable. Whether generation, retrieval, and checking stayed inside the student is.

Hiroko Warshauer mapped the adult moves that keep demand high while helping a stuck student: probe what they have already tried, set two student methods side by side, ask them to explain a stuck peer's idea without handing over the path. That is a hint ladder in human form. A tool can be asked to do a crude version — request the attempt, name a common mistake, withhold the last step. Consumer chat, by default, skips those moves. A hint tutor with teacher-written mistakes in the prompt is the crude version of the probe.

Louis Deslauriers and colleagues documented an illusion that will show up at your table: students in a passive lecture felt they had learned more than students in an active classroom, while learning less. Fluency of the hour is a feeling. The unaided ticket is the measure. If the only thing you grade is the session artefact, you have hired the illusion as a helper.

Cognitive offloading is old and not a moral failing. Risko and Gilbert defined it as using the environment to reduce internal processing. Calculators and maps produce the same trade-off: performance up, unaided skill down, if the offloaded piece *was* the learning target. Generative tools extend the offload from memory and calculation to drafting, reasoning, and evaluation — the work school exists to form. The question is which part of the work still has to be produced by the student.

Wrong answers you should be able to hear

1. *"I don't get it" — with a blank page.* Often the student has not yet attempted. The next move is a smaller first step, not a lecture and not a finished solution. Try: "Write one thing you know that might matter. Then we look."
2. *A perfect, fluent solution in a voice that is not theirs, produced in minutes.* Rescue already happened, by you or by a tool. The next move is an unaided twin problem, same idea, different numbers or a different source. Grade that one.
3. *Copied arithmetic or a copied claim the student could have caught.* The evaluative loop was skipped. Ask them to check one line out loud, with the tool closed. If they cannot, the hour was performance, not learning.
4. *Instant polish on a first draft of writing.* The artefact improved; the knowledge may not have. Fan and colleagues found that ChatGPT users wrote

better essays and showed no significant knowledge or transfer gains, with fewer evaluation moves in the process. Ask for a restatement in their words, then a short unaided write.

5. *Frozen silence for a long stretch, or tears, or random clicking.* That is unproductive struggle: overload, a goal that is unclear, or a task that does not fit what they know yet. Shrink the task. Show a worked example. Sit beside. Rescue of the *person* is not rescue of the *problem*.

Five-minute parent warm-up

Before the lesson, take one problem from today's work and solve it yourself on paper. Circle the step you would be tempted to narrate too soon. Write one hint that names a move without naming the answer — "What has to be the same on both sides?" or "Quote the line that the question is actually about." Hide the finished solution. That hidden page is your answer key. The student does not see it until they have attempted.

If you feel rusty on today's idea, use those five minutes to become fluent enough to hear a wrong turn. You are not taking over the pencil. You are becoming a person who can ask a good question. Many adults feel rusty. That is normal. One worked problem in your own handwriting is enough for this hour.

How to teach it this week

Session shape, forty-five to sixty minutes. Same bones every subject. The wording below is math-flavored so you can say it; swap the object and keep the moves.

Warm-up (5 minutes). Retrieval of something already known, unaided, short. Two facts, one prior problem, or one sentence from yesterday's reading. No tool. The point is to start the hour inside the student's head.

Say: "Two from last time, on paper. I am not helping yet. Then we look."

Short model (5–8 minutes). For a *new* idea, you show one worked example. Think aloud the decision, not the autobiography of your anxiety. Point to the structure: what is given, what is asked, what must stay equal, which line is a definition and which line is a guess.

Say: "Watch this one. I will write every step. Your job is to tell me, at the end, which step you would try first on a new problem."

If the idea is *not* new, skip the full model. Ask them to generate first. Expertise reversal has already moved the slider.

Student attempt (10–15 minutes). A near-twin problem, tool closed, pencil in their hand. You sit close enough to see, far enough not to take over.

Say: “Your turn. Attempt first. If you are stuck after three honest minutes, tell me what you tried, and I will give one hint — not the answer.”

Then wait. Waiting is the lesson. Count a slow thirty after they stall before you speak. If you grab the pencil, you have converted demand into procedure.

One good question (2 minutes). Not “do you get it?” They will say yes. Ask a question that forces generation.

Say: “Why that step, not the one before it?” or “If I change this number, what happens to your answer?” or “Point to the sentence in the source that your claim actually rests on.”

Mixed practice (10–15 minutes). Two more problems, not clones of each other. One is today’s move. One is yesterday’s move in disguise. Mixing feels harder and teaches *when* to use the move, not only *how*. If they fly through blocked copies, they have memorized a shape, not a decision.

Exit ticket (5 minutes). One unaided item, tool closed, collected. This is the measure. The practice pages are not.

First problem for the student. Pick the smallest version of today’s idea that still requires a decision. In algebra, one linear equation with a visible like-term, not a word-problem tower. In writing, three sentences that must include a claim and one piece of evidence, not a five-paragraph essay. In history, two short sources and one question: who is speaking, and what can they not know? The first problem should be finishable after a struggle. If it is not, you have assigned unproductive struggle.

Exact wording when they ask for the answer.

- “Show me the last line you are sure of.”
- “What have you already ruled out?”
- “I will give a hint in one sentence. Then you go again.”
- “I am not taking the pencil.”

How to fade help. Day one, a full worked example and a twin. Day two, a faded

example with one blank step. Day three, attempt first, then a hint from a small list you wrote: a common mistake, a definition, a pointer to a diagram. Day four, unaided. If they succeed unaided, stop showing full solutions. If they fail unaided, step back one rung — faded example, not a lecture.

When to stop talking. After the hint. After they resume writing. After the exit ticket is in their words. Narrating while they work is rescue with more sentences. Your job in the attempt window is to notice whether the struggle is productive (effort toward sense-making) or unproductive (overload, lost goal, wrong task), and to change the task size — not to finish the item.

Illustration, not a reported family: a ninth-grader pastes a quadratic word problem into an unguarded chatbot, copies a worked solution that is wrong in the arithmetic, and the homework looks strong. The unaided twin, given the next day, is the hour that tells the truth. The difference is not the model. It is whether the generative work stayed inside the student.

A hint ladder you can say this week. Write it on a card. Climb only as far as needed.

1. “What is the question asking, in your words?”
2. “What do you already know that might matter?”
3. “Show me the last line you trust.”
4. Name a common mistake for this idea — “People add the denominators. Is that a size, or a count of pieces?”
5. Give a faded step: the first line filled in, the rest blank.
6. Only then, model one line, and hand the pencil back.

For writing, the same ladder is: restatement of the prompt; one claim in their words; one piece of evidence they already have; then a question about order; then, if needed, a sample sentence *about a different topic* so they still have to write today’s idea. For history: who wrote this, when, what they can see, what they cannot know; then a second source; then, if needed, a sentence stem, not a paragraph.

Human high-dosage tutoring remains the benchmark against which software should be read. Nickow, Oreopoulos, and Quan pooled experimental preK–12 tutoring near three-tenths of a standard deviation. Earlier computer-assisted practice typically sits lower. Every positive generative-AI field result in Chapter 3 either sat inside a teacher-run session or helped the tutor. The adult in the room is not decoration. The adult sets the struggle policy and notices when struggle turns unproductive.

Practice that actually builds learning

Blocked practice is for a *new* move. Several of the same kind in a row, while the idea is still forming. The student should be able to say what the move is.

Mixed practice is for *when* to use it. After the move exists, put today's item next to last week's item and next to a near-miss that looks similar and is not. Mixing is a desirable difficulty. It will look worse in the session and hold better next week.

Retrieval is for facts already known. Two minutes, unaided, at the start or the end: definitions, number facts they can already get right untimed, a date, a quote, a theorem name. Retrieval is a learning event, not only a quiz. For a beginner on a dense, highly connected idea — a first proof, a first causal argument in history — unassisted generation can overload working memory. A worked example or a hint sequence is the better first move. Retrieval comes after a schema has begun to form.

One incorrect example to diagnose. Put a finished solution on the table that contains a single, typical error. Ask the student to find it and to say why it is an error. This is cheaper than another worksheet, and it trains the checking loop that a fluent model will otherwise skip.

Kitchen, money, and making can motivate, or give a second picture. They do not replace the number line, the written equation, the unaided paragraph, or the source on the page. A recipe that “uses fractions” is not a substitute for $1/2 + 1/3$ on a line.

Subject notes, short, so the week has somewhere to stand.

Mathematics. Hint ladders help; answer-giving chat harms subsequent unaided work in the school study Chapter 3 already taught. Access to a tutor app is not tutoring. In one large middle-school trial, most students barely messaged the tutor, and the effect looked like extra practice without the chatbot. Be in the room. Require an attempt. Delay the first hint. Ground hints in the mistakes *this* course actually makes.

Writing. Using a tool to explain a concept, after the student has a draft, is a different hour from using a tool to generate the first draft. Contractor and Reyes found delayed unaided essay gains for the first pattern, not the second. Fan's study is the artefact warning: better essays, no significant knowledge or transfer difference, thinner evaluation in the process. Grade a process log and an unaided write, not

only polish.

Science. A custom tutor with the teacher’s steps in the prompt, on first-pass conceptual material, can raise an immediate post-test. Open inquiry with a general chatbot is a different design: Stadler, Bannert, and Sailer found lower load and worse scientific argument than search. Structured tutoring is not open inquiry. Lab work stays in the hands and eyes. An overnight lab report is rescue.

Reading and history. Direct generative-AI trials on learning to read are thin compared with math. That thinness is itself information. A summary of a chapter the student has not read hides the work. Listening to a model explain the chapter is not reading. If the objective is decoding, a rewrite-on-demand tool has not differentiated instruction; it has removed the grapheme-phoneme work. In history, the student reads sources first. A confident machine narrative is then a problem to mark: sourced, inferred, or fabricated. “Write the DBQ with the chatbot” grades the model.

Computer science. Autocomplete is a worked example at keystroke speed. Novices shepherd (steer suggestions instead of writing) or stray (accept a wrong suggestion into a debug hole). Benefits concentrate among students who already have stronger monitoring skills. Manual tracing first. “Explain this generated function.” “Write this without the model.” Agentic repo-builders that plan and submit the project are a different offloading regime. Stanford’s CS221 course, in Spring 2026, treated that refusal as legitimate on the assignment that forms the skill.

A sample week, labelled as a teaching sequence rather than a district calendar. Monday: new idea, worked example, twin, unaided ticket. Tuesday: faded example, mixed with yesterday’s idea, unaided ticket. Wednesday: attempt first, one hint from your card, mixed practice. Thursday: incorrect example to diagnose, then two mixed items. Friday: unaided quiz on the week’s move, tool closed, five minutes oral — “teach it back.” If Friday fails, next week is not a new chapter. Next week is the faded example again.

Illustration, not a reported family: an intro programming student accepts a string of autocomplete suggestions, the program runs, and they cannot explain a loop. That is shepherding. Another student accepts a wrong suggestion, spends the hour in a hole they did not dig, and concludes they are bad at the subject. That is straying. The course that requires tracing, an explanation of a generated function, and a write-without-the-model item is the slider set the right way.

For the student

This idea is simple, and it is not easy. You learn the things you actually do. If a tool does the hard part, the hard part does not become yours.

Productive struggle is the name for the good kind of stuck. You can see the goal. You have a way to start. The next step is not obvious. You try anyway. That effort is not a sign that you are bad at the subject. It is the subject happening.

Unproductive struggle is different. You cannot tell what the question is asking. Every path feels like noise. You have been stuck long enough that you are no longer thinking. That is the moment to ask for a smaller step, a worked example, or a hint — still not the finished answer.

Tiny worked example. Suppose the problem is: a number plus four is ten. What is the number?

A rescued version looks like this: someone (or a chatbot) writes $x + 4 = 10$, so $x = 6$, and you copy it. You have a correct homework line. You have not decided anything.

A struggled version looks like this: you write *something* $+ 4 = 10$. You try 5, get 9, which is too small. You try 6, get 10. You write 6. Then you check: $6 + 4$ is 10. The check is yours.

The algebra later is the same move with a better notation. The check remains yours.

Two tries.

1. Close every chat window. Take one problem from today's work. Set a timer for five minutes. Write every attempt, including the wrong ones. If you finish early, write one sentence: "I knew this step because..."
2. Still without a tool, take a second problem that looks related but is not a copy. If you get stuck, write the last line you are sure of, then ask your parent for *one* hint. Use the hint. Finish on your own.

Explain it back. In four sentences, tell a person what "attempt first" means, why a fluent copied answer can hide a mistake, and what you will do this week when you feel stuck. If you cannot explain it, you are not ready to use a helper tool on this idea.

Challenge. Your parent puts a finished solution on the table with one error in it. Find the error. Say what the person (or the model) seemed to think. Write the correction. This is judgment practice. Chapter 6 will go further. You can start now.

A writing version of the same idea. Suppose the assignment is three sentences: a claim about a short story, and one piece of evidence. A rescued version is a chatbot paragraph that sounds like a review. You paste it. The page looks adult. You cannot point to the line in the story that the claim rests on. A struggled version is slower: you copy the line from the story first, then write what you think it means, then write the claim. If a tool is allowed later, it may ask you questions about *your* sentences. It does not get to write the first ones.

You are allowed to struggle. You are allowed to ask for a hint after an honest attempt. You are not required to look fluent in the first ten minutes. Fluency of the session is a poor measure of the skill.

If it isn't clicking

Three diagnostics. No shame. The next move is a task change, not a character speech.

1. They attempt, then stall on a step they do not have a name for. The idea is new and heavy. Working memory is full. Slow down. Show one worked example. Then a faded example with one blank. Then a twin. Save mixed practice for after they can complete a twin. A human tutor is worth considering if this pattern lasts more than two weeks on the same idea and the student is losing the thread of the course.

2. They refuse to attempt, or they paste first. The culture of the hour is rescue. Change the rule in writing, on the assignment: "Attempt visible. Hint after three minutes. Tool closed for the exit ticket." Sit through the silence once without filling it. If a tool is in the house, it waits until after the attempt. If they still paste, the unaided exit ticket is the grade for this skill. You are not punishing curiosity. You are measuring the job.

3. They attempt hard, the hints do not help, and frustration spikes. The task is the wrong size, or a prior skill is missing. Go back one idea — the fraction work under the algebra, the decoding under the chapter questions, the definitions

under the essay. Productive struggle requires a task in reach. If the student cannot succeed with effort and a hint, the difficulty is not desirable. Shrink it. If you cannot find the missing prior skill, get a human tutor for that skill. A chatbot that keeps talking is not a diagnosis.

When to go ahead: unaided exit tickets are mostly right, the student can explain the move, and mixed practice is only a little slower than blocked. When to slow down: the unaided ticket fails, or the explanation is a recitation of your hint. When to get a human tutor: the missing prior skill is a whole band behind, or the hour has become daily conflict. A person who can sit beside the work still outperforms, on current evidence, an unsupervised window.

If two of the diagnostics are true in the same week, treat the hour as too large. Cut the session to a worked example plus one twin plus an exit ticket. That is still a complete lesson. A long session of unproductive struggle teaches the student that this subject is a place where they drown. A short session of managed struggle teaches the next move.

Tools, including AI

Optional helpers for the adult. The student attempts first. You hold the answer key. A tool can explain the *idea to you*, make extra practice with answers hidden, write a hint script, or help you diagnose work already done. It does not sit open during the attempt.

Use a tool this week in one of four ways:

- Ask it to explain today's concept to *you*, then close it and teach from paper.
- Ask it for three extra problems of the same type, answers at the bottom, which you cover.
- Ask it for a hint ladder: attempt prompt, then a small hint, then a bigger hint, never a full solution on the first request.
- Paste a student's *already finished* wrong work (no name attached) and ask what misconception it might show. You still decide.

A hint-only tutor with an attempt gate is the design that matches this chapter. Unguarded chat that returns a full solution is the design Chapter 3 already showed can raise practice and weaken the unaided exam. Adaptive systems that sequence items from a learner model — the older practice engines many families already

own — are not chat and are not agents. They can run retrieval. They cannot replace the unaided ticket.

Keep this box short on purpose. The struggle policy is the chapter. The tool is optional furniture.

Rules for this week

- Child attempts first.
- Parent holds the answer key.
- Hints before answers.
- Tool closed for warm-up and exit ticket.
- No photo of the worksheet into a chatbot for “what’s the answer.”
- No unsupervised window during the attempt.

If you use a school-issued Socratic tutor, stay in the room. Access is not tutoring. Engagement is still your job.

What “done enough” looks like

Move on when most of this is true, on more than one day, without you holding the pencil.

- The student attempts before asking.
- A hint is enough more often than a full model is.
- The unaided exit ticket is mostly correct, or the error is a small slip the student can catch when asked to check.
- The student can explain the move in their own words.
- Mixed practice is slower than blocked practice, not a collapse.
- You have faded a full worked example to a blank page on this idea.
- A helper tool, if used at all, waits until after the attempt and does not produce the graded artefact.

Placement is by skill, not by birthday. A “grade 8” book is a publisher’s scope. If fractions as numbers are not yet solid, that is the work, even if the cover says algebra. If the student can already generate and check, stop showing full solutions; the slider has moved.

Done enough is not a fluent homework produced in four minutes. Done enough is a skill that survives when the window is closed.

Chapter 8 will put this into two lanes of assessment: formation work graded unaided, and later AI-world performance graded as judgment, logged. This chapter only needs the first lane to be real. If the formation lane is empty — if every page can be produced by a model — the later lane has nothing to stand on.

Keep the struggle in the student. Ask, wait, hint, then model. The next chapter is why that student still needs knowledge in the subject: judgment has to have something to judge.

Chapter 6

You cannot judge what you do not know



Figure 7: Marked pages, a magnifying glass, old books, and a closed laptop.

Why this matters

A fluent paragraph can look like thinking. It has a claim, a because, a however, and a neat ending. If the student does not know the subject, that paragraph is a costume. They cannot tell whether it is right. They can only tell whether it sounds finished.

Daniel Willingham wrote the constraint this chapter lives on, in plain teacher language, in 2007. People have treated critical thinking as a skill like riding a bicycle: learn it once, take it anywhere. Thinking is not that kind of skill. The processes of thought are tied to the content of thought. You can teach maxims — consider another side, look for evidence — and without background knowledge and practice the student will not be able to use them. Historical thinking, scientific thinking, and literary thinking are partly distinct. Strategies help at the margin. Knowledge does the work.

This is worth the week because the tools in the house are extremely good at the *surface* of judgment. They will interpret, analyze, evaluate, infer, and even sound self-correcting on a topic the student has never studied. If the student lacks the schema to check a claim, evaluation is a coin flip in a nice voice.

Chapter 5 kept struggle in the student so a schema can form. This chapter is what that schema is for: something to judge. An “AI literacy” unit that recites maxims while the model writes the history essay is practicing the habit of outsourcing the domain and reciting the maxim. That is the opposite of the skill.

You do not need a philosophy degree to teach this. You need today’s subject well enough to hear a fluent error, and a student who has actually read, solved, or observed the thing they are asked to judge.

For the parent: understand it yourself

Here is the everyday picture. Someone asks you to decide whether a used car is a fair price. If you know what a timing belt is, what rust at that seam means, and what that engine should sound like, you can judge the seller’s story. If you do not, you are judging confidence. A smooth talker and a smooth paragraph are the same trap. Critical thinking, in the bicycle version, says: “consider the source.” That maxim does not help you if you cannot tell a belt from a hose.

Here is the precise picture. Willingham's point is not that transfer never happens. It is that transfer is not the default, that it is bound to a domain, and that teaching the maxim without the domain is a habit with a weak record. A student who has learned to consider alternative explanations in a unit on vaccines does not automatically do so in a unit on Reconstruction, unless the Reconstruction knowledge is there to generate the alternatives.

Two older checklists are still useful *inside* a subject. Robert Ennis described critical thinking as reasonable, reflective thinking about what to believe or do, plus dispositions (seek reasons, stay open, take a position when the evidence warrants) and abilities (clarify, analyze arguments, evaluate sources, infer). Peter Facione's account, which still sits under widely used tests, names six skills: interpretation, analysis, evaluation, inference, explanation, and self-regulation, plus habits of mind. Those lists help you design a task. They do not tell you what the student must already know in order to execute any item on them. Willingham is the constraint that keeps the lists honest.

Default chatbot use collapses evaluation into a single move: keep this paragraph, or ask again. Students then practice neither source evaluation nor argument construction. They practice a binary. Fan and colleagues watched the writing version: better essays, no significant knowledge or transfer gains, fewer evaluation and orientation moves than students who talked with a human expert. Stadler, Banert, and Sailer watched the science version: lower load, worse reasoning than search. Bastani's unguarded math tutor, in a probe, was correct only about half the time; students still copied, including arithmetic they had the school math to catch. Prather and colleagues watched novices with code autocomplete shepherd or stray; students who already monitored their own thinking benefited, and students who did not were harmed.

None of that is a law that "AI makes you worse at thinking." Contact is not the variable. Whether the student still has to produce the judgment is.

Literacy frameworks — ISTE's six roles, AI4K12's charts, UNESCO's competencies, Cornell's incoming-student modules, a Department of Education grant definition — are usable when they sit on domain knowledge. They are checklists students can recap and not execute when they do not. Cornell's program asks a better question than "can you prompt": does this use *contribute to learning*? That question is Willingham in administrative clothing. This chapter is how you ask it at the table.

There is not yet a randomized trial showing that a K–12 “evaluate the chatbot” curriculum transfers critical thinking to tasks with no chatbot. That hole is real. It is not a reason to skip domain-loaded practice. It is a reason not to treat a completed literacy module as a formed skill.

Wikipedia is the comparison many families reach for, and it is the wrong comparison for fluency. A weak encyclopedia page still has seams: a history tab, a talk page, a citation you can open. A model paragraph has the grammar of a completed judgment and no history tab. Students trained to distrust an unsourced wiki sentence have not been trained to distrust a paragraph that sounds like their teacher. Fan’s ChatGPT group did not fail because the essays looked machine-made. The essays scored well. Knowledge and transfer did not move. Stadler’s users did not report confusion. They reported lower load and wrote worse arguments. Comfort is the pedagogical problem. Comfort is not a tell.

A 2026 distinction between dependent offloading and autonomous offloading is useful even before anyone treats it as a settled test: how often the student uses a tool is not the variable; whether they kept the evaluative loop is. Default chat does not keep it. Guarded, domain-loaded tasks can.

Wrong answers you should be able to hear

1. *“The AI said it, so it’s probably right.”* The student is judging fluency. Ask for a restatement in their words, then one check against a source, an equation, or a measurement they already have.
2. *A perfect “both sides” paragraph on a topic they have not studied.* The maxim was recited. The domain is missing. Send them back to the reading, the lab, or the problem set. The critique comes after.
3. *They mark grammar and tone in a model essay and miss a false fact.* They have evaluated style, not warrant. Give a rubric with a line for “what would make this false,” and require one mark on that line.
4. *They cannot quote or restate their own AI-assisted essay.* Encoding failed. Kosmyna’s small preprint found a version of this in a lab; you do not need EEG. Close the tool and ask, “Read me your second paragraph.” If they cannot, that paragraph was never theirs.
5. *“I did the AI-literacy unit, so I can use any tool on any assignment.”* Literacy was treated as a bicycle. Name the subject knowledge this assignment requires, and check it unaided before any tool is allowed.

Five-minute parent warm-up

Take one claim from today's work — a sentence in the textbook, a line the student might paste, or a claim you generate yourself. Without a tool, write: (1) the claim in your own words, (2) one fact you would need in order to judge it, (3) one way it could be wrong. That is interpretation, knowledge, and evaluation on a sticky note. If you cannot do those three, you are not ready to hear the student's critique yet. Spend the five minutes on the fact, not on a speech about critical thinking.

If today's subject makes you feel rusty, choose a smaller claim. You need enough of the idea to hear a wrong turn, not enough to guest-lecture. One honest sticky note beats a printed rubric you cannot apply.

How to teach it this week

Session shape, forty-five to sixty minutes. The object of the hour is a *claim in a subject the student has already worked*. New content belongs in the struggle hour of Chapter 5. This hour judges something they have a chance of knowing.

Warm-up (5 minutes). Unaided retrieval of the domain. Two facts, one definition, one date, one equation, one quoted line from the source they read. Tool closed.

Say: "Before we look at anyone else's paragraph, tell me three things you already know about this. Paper only."

Short model (5–8 minutes). You take a short, imperfect claim — one you wrote, or a model paragraph you printed and marked in advance — and think aloud *only* the judgment moves.

Say: "Watch what I do, not how smooth I sound. First I restate. Then I mark what is a fact, what is an inference, and what I cannot check. Then I say what would make this false."

Keep the model under a dozen sentences. A long performance teaches performance.

Student attempt (10–15 minutes). They get a new claim in the same domain, tool closed at first. They must produce: a restatement, a mark-up (fact / inference / unchecked), and one check against a source or a calculation they already have.

Say: "Your turn. Restate without looking. Then open the source, not a chatbot."

Mark three places. You are grading the claim, not rewriting it.”

One good question (2 minutes).

Say: “What would you need to know that you do not know yet in order to trust this?” If they cannot name a gap, they have not evaluated. They have admired.

Mixed practice (10–15 minutes). One claim that is mostly true with a quiet error. One claim that is fluent and hollow. One claim from last week’s topic, so they have to notice the domain has changed. Mixing here is “when is this a fact, an inference, or a costume?”

Exit ticket (5 minutes). Oral, one minute, tool closed: “Restate the claim. Say one thing that supports it. Say one thing that would make it false.” You are listening for knowledge, not for the word *however*.

First problem for the student. A paragraph of eight to twelve sentences on material from the last two weeks, with one invented detail and one leap. Math: a worked solution with a correct method and an arithmetic slip. History: a narrative that names a real event and attributes a motive no source gave. Science: a mechanism that explains the classroom demo and also “explains” a result the demo never showed. Writing: a thesis that sounds wise and does not match the passage. The first problem should be catchable by a student who did the reading or the problem set. If it is not, you have assigned theater.

Exact wording you can say.

- “Put that sentence in your mouth without looking at it.”
- “Where is that in the source? Finger on the line.”
- “Is this a fact we have, a guess, or a fluent gap?”
- “Give me a counterexample. One is enough.”
- “Why did you keep this sentence and cut that one?”

How to fade help. Early in a topic, you supply the claim and the source packet, and you sit through the restatement. Later, they find the claim (including a model output) and run the same marks. Later still, they generate a claim first, then may use a logged tool as a critic, then defend the keep-and-cut decisions aloud. Help fades from “I will show you what a mark-up looks like” to “you run the loop; I hear the defense.”

When to stop talking. After they have restated. After they have put a finger on a line. After the oral ticket. If you narrate the critique, you have done Facione’s

evaluation *for* them. The model is already eager to do that. The hour exists so a human still has to.

Map the six skills onto tasks, not posters.

- Interpretation: restate the model's claim in the student's words before anyone uses it.
- Analysis: diagram the argument; mark premises left implicit.
- Evaluation: score a model essay against your rubric; find the weakest warrant.
- Inference: given the conclusion, generate a counterexample.
- Explanation: oral defense — why keep this, cut that.
- Self-regulation: a process log — where they accepted, challenged, or ignored the model.

The student must interpret, check, and orally defend. If any of those three is missing, the loop has been outsourced no matter what the syllabus says about literacy.

Illustration, not a reported family: a tenth-grader uploads a chapter, asks for a summary, and never reads the chapter. The notes look complete. Decoding, fluency, and the knowledge that later reading needs were the learning objectives. The off-loaded process *was* the skill. Frequency of use is not the diagnosis. The missing evaluative loop is.

Walk restatement first, every hour. If they cannot put the sentence in their own words, they do not have a claim to evaluate. They have accepted a texture. Restatement is retrieval. It is also the cheapest test of “could you quote your own essay,” and it needs no lab.

Counterexamples next. Given the model's conclusion, they produce a case that does not fit. If they cannot, they have admired the conclusion. In mathematics this is a failed identity. In history it is a source that will not sit still in the narrative. In science it is an observation the mechanism cannot explain. Generation of the counterexample is the inference skill Facione named, and a model will happily generate one *for* them if you let it. The assignment has to require theirs.

Practice that actually builds learning

Blocked practice, this week, is several mark-ups in the *same* domain: three model claims about the same chapter, the same lab, or the same type of equation. The

student should get faster at restatement and at finding the local error.

Mixed practice is claims from two domains in one sitting, plus one “style-only” distractor that is grammatically fine and factually empty. The student should have to choose which knowledge to bring.

Retrieval is the knowledge the judgment sits on. Five facts from the unit, unaided, before any critique. A critique hour with no retrieval is a literacy costume.

One incorrect example to diagnose: a fluent paragraph that uses the right names and the wrong causal arrow. Ask what the writer would have to believe. Ask what source would break it. This trains evaluation, not hunting for a missing comma.

Domain-loaded versions, so the week is not an abstract “critical thinking” worksheet.

History. Historical thinking is sourcing, corroboration, and contextualization. A large language model that writes a confident narrative is a primary-source problem: it looks like a secondary source and behaves like an unsourced claim. Sequence: student reads the sources first; model (or you) produces a narrative; student marks sourced, inferred, fabricated; oral defense. There is almost no subject-specific generative-AI randomized trial in history as of the research behind this book. You still refuse to hide the work. Waiting for a trial before requiring sources is how the work gets hidden.

Science. Structured tutoring on a known objective is not open inquiry with a general chatbot. Kestin and colleagues showed a custom physics tutor with instructor-written steps can beat even a strong active classroom on an immediate post-test; delayed retention was not measured, and the lessons were not “create a theory.” Stadler’s study is the inquiry warning: ease, thinner argument. For a known mechanism, a hint tutor plus an unaided check. For “what should we conclude,” the student searches, reads, and argues; a chatbot that lowers load is the dulling design already measured.

Mathematics. Judgment is catching the step that does not follow. Print a solution. Ask which line is a definition, which line is algebra, and which line is a wish. If they cannot find a copied arithmetic error they have the skill to catch, they did not judge. They copied. Chapter 3 is the school study; this hour is the habit.

Writing. Contractor and Reyes split *augmentation* (use the tool to explain) from *automation* (use the tool to generate text). Delayed unaided essay gains showed up for the first, not the second. Practice: student drafts first; tool may question

the draft; student keeps a log of accepts and refusals; a short unaided write or oral defense is part of the grade. Fan's missing evaluation cycles are what this practice puts back.

Computer science. "Explain this generated function. Trace it. Write the next one without the model." UC San Diego's introductory sequence, in a company customer story, put a manual-competency gate before Copilot. Steal the gate. The story is not a trial of cloud agents. Stanford's refusal of agentic repo-builders on the project is the same idea at the next product class.

Hallucination-checking is source evaluation under high fluency. It is harder than evaluating a clumsy human paragraph, and therefore useful *if* the student has enough knowledge to notice. The Nigeria after-school study taught students that Copilot hallucinates and that they remain responsible, inside a teacher-run lab with science-of-learning prompts. The English outcome moved. The treatment was not the chatbot alone. Teach checking as a responsibility on material they have studied. A "spot the fake news" worksheet on a topic they have not learned is the bicycle again.

ISTE's six roles — Learner, Researcher, Synthesizer, Problem Solver, Connector, Storyteller — pass Willingham's test only when each role sits on a domain. Researcher without documents is a pose. Synthesizer without a "then verify" clause is the model synthesizing. Storyteller who begins with the model's draft has automated the voice. Use the names if they help you talk. Grade the deeds.

A conversation scored as "analysis" on a chat log can be the model analyzing. Willingham's test is what the student can do when the log is closed. Treat log-scored "critical thinking gains" as a description of the chat, not as a formed skill in the subject.

Singapore sequences literacy by age rather than by brochure: print-first in the early primary years, ministry-designed features under teacher supervision later, then increasing student ownership with a duty to evaluate output, check accuracy, and acknowledge use. That is a state choosing when the student is the judge. Steal the sequence, not a product. AI4K12's charts run the same way if you let them: about AI first, then with constrained tools, with "evaluate fairness" in high school sitting on enough mathematics and civics to inspect a system. A fourth-grade chatbot conversation about topics the child cannot yet read fails the same test this chapter already gave you.

For the student

Judgment is not a vibe, and it is not a personality. It is what you do with what you know.

If you do not know the subject, you cannot tell a fluent error from a fluent truth. That is not an insult. It is how minds work. You would not want a person who has never cooked to “critically evaluate” a recipe by how nice the sentences are. School subjects are the same.

A chatbot is very good at sounding like a person who already judged. Your job is to remain the person who actually does.

Tiny worked example. Claim: “Water boils at 80°C at sea level.”

Restatement, in your words: the temperature where water turns to steam, at the ocean’s height, is eighty degrees.

Check against what you know: you have seen water boil. You have been taught 100°C at standard pressure. Eighty is wrong, or it is about a different altitude, which the claim did not say.

Counterexample: a kettle at home, at sea level, with a thermometer, if you have one; or the definition in last week’s notes.

You did not need a maxim about “multiple perspectives.” You needed a fact.

Two tries.

1. Close the chat. Take one paragraph from this week’s reading or one worked problem. Write the main claim in your words. Under it, write one fact that supports it and one fact that would break it.
2. Your parent gives you a short machine-written paragraph on the same topic. Mark each sentence S (in the source), I (inferred), or F (fabricated or unchecked). Defend one mark out loud.

Explain it back. Tell someone, in five sentences, why knowing the subject comes before “critical thinking” about a chatbot, and what you will do this week before you accept a fluent paragraph. If your explanation is only “you have to fact-check,” add *what with*.

Challenge. Given a model conclusion in your current unit, invent a counterexample the model did not consider. In math, a case where the identity fails. In history,

a source that does not fit the story. In science, an observation the mechanism cannot explain. If you cannot invent one, you do not yet know enough to judge. That is a study assignment, not a character flaw.

Here is a second tiny picture, this time from history. A paragraph says, with great confidence, that a famous speech “united the country the next day.” You have read two newspapers from the week after. One praises the speech. One ignores it. Your job is not to write a smoother paragraph. Your job is to mark “united the country” as a leap, point to the two papers, and say what you still cannot know — who was silent, who was not in the papers at all. The chatbot will happily give you the leap. The sources give you the work.

You are allowed to say “I do not know enough yet.” That sentence is judgment. “It sounded good” is not.

If it isn’t clicking

1. They recite maxims and still accept fluent errors. The bicycle is in charge. Slow down. Put the literacy language away for a week. Run retrieval on the unit until they can produce five facts unaided. Then give a one-error paragraph that those five facts can catch. If they catch it, the knowledge was the missing piece. If they still miss it, model the mark-up once, then give a twin.

2. They can find typos and cannot find falsehoods. They are grading style. Change the rubric in front of them: one line is “what would make this false,” and that line is required. Give paragraphs that are grammatically clean. Praise the caught fact, not the caught comma. If this lasts, the reading or the problem set is not actually being done. Go back to Chapter 5’s attempt rule on the source itself.

3. They freeze, or they say everything is biased, or they refuse to take a position. Open-mindedness without knowledge becomes fog. Shrink the task: one claim, one source, one keep-or-cut. Ennis includes taking a position when the evidence warrants. Practice that on small, checkable items before anyone asks them to “evaluate AI” in the large. If anxiety is the real block — they fear being wrong in public — make the first defenses private, on paper, to you only. A human tutor helps when the freeze is bigger than one unit: a reader who cannot decode cannot judge a history narrative, and a student without fraction sense cannot judge a proportion argument.

When to go ahead: they restate without looking, they catch a planted error in a domain they have studied, and they can say what they still do not know. When to slow down: the oral ticket is a recitation of your critique, or they cannot retrieve the unit facts. When to get a human tutor: the domain gap is a whole band, or judgment hour has become a daily argument about tools instead of a five-minute mark-up.

If the student is quick to call everything “biased” and slow to check a date, they have borrowed a grown-up word as a shield. Bias talk is part of later civic literacy. It is not a substitute for “is this number the one in the table.” Bring them back to a checkable item. Confidence in a small, correct check is the path to later, larger judgment. Fog is not sophistication.

Tools, including AI

Optional helpers for the adult. The student still has to produce the judgment. You hold the source packet and the answer key.

A tool this week can:

- Draft a *wrong* paragraph on last week’s unit for you to print and mark up in advance. You check every fact before the student sees it.
- Write three claims at different quality levels on the same topic, answers hidden, for mixed practice.
- Ask *you* what a typical misconception in this unit is, so you can plant one.
- After the student’s own draft exists, generate questions about *that* draft. The student answers. The tool does not rewrite.

A tool this week is a poor choice as:

- The teacher of “critical thinking” in the abstract.
- The first drafter of the paragraph being graded as the student’s judgment.
- An unsupervised conversation about a topic the student has not studied.

Mollick and Mollick’s menu — tutor, coach, mentor, teammate, student, simulator, tool — is a prompt library, not a theory of thinking. The Nigeria study actually used science-of-learning prompts inside a teacher-run lab. The teacher still chose the role. The one role that forces explanation is the model as *student*: the human has to teach it. That only works if the human knows the content. Coach and mentor

can keep the loop if they refuse the product and demand an attempt first. Teammate and tool will finish the job if the assignment lets them.

Rules for this week

- Knowledge first: retrieval before critique.
- Child restates before anyone uses a model sentence.
- Parent holds the source or the key.
- Oral defense on anything a model touched.
- No unsupervised chatbot as a stand-in for the reading.

What “done enough” looks like

Move on when most of this is true on material from the current unit, not on a general “media literacy” worksheet.

- The student restates a claim without looking.
- They catch at least one planted error that requires unit knowledge, not just spelling.
- They can point to a source, an equation, or a measurement for a keep-or-cut decision.
- They can name one thing they still do not know that would change the judgment.
- A one-minute oral defense sounds like their voice and their facts.
- A process log, if a tool was used, shows at least one challenge or refusal, not only accepts.
- They no longer treat an AI-literacy badge, a maxim, or a smooth paragraph as proof of thought.

Placement is by knowledge in the subject, not by age, and not by how sophisticated the tool talk sounds. A twelve-year-old who knows the chapter can judge a paragraph about the chapter. A seventeen-year-old who can name six ISTE roles and cannot catch a fluent error in the course they are taking is not ready for the second assessment lane in Chapter 8.

Done enough is not “they know to be skeptical of AI.” Done enough is: they know enough *in this subject* to catch a fluent error, and they will do the restatement, the check, and the spoken defense.

Willingham’s constraint does not forbid literacy. It forbids treating literacy as a

replacement for history, mathematics, science, and the rest of the curriculum a model is willing to write. Keep building the something that judgment needs. The next chapter is who hosts the tools, and who is still allowed to turn them off.

Chapter 7

Who hosts, who turns it off



Figure 8: A closed book, a school bell, and a dark computer screen on a desk.

Why this matters

A tool in a child's life needs a host: someone who can see the log, receive a safety ping, delete the memory, and switch the thing off. If no one can do those jobs, you do not have a school tool. You have a roommate.

This week's idea is practical. Default host is the institution — a district, a co-op, an umbrella school, a ministry platform, or *you* as the school of record — under rules you can name. A parent-installed autonomous agent on a teenager's laptop is a home tutor at most. It is not a child-safety school product. Grok Bot, launched in August 2026, has no education SKU: its own pages illustrate sales, inboxes, expenses, and bug reproduction, and they do not mention schools, students, or teachers.

The sentence you will hear about labor is folklore, then we teach: “teachers who use AI will replace those who don't” is a recycled slogan, not U.S. Department of Education, NEA, or AFT doctrine. What the field studies actually show is orchestration. A grown-up chooses the lane, sets the struggle policy, stays in the room, hears the oral check, and keeps the off-switch.

This is worth the week because hosting is how Chapters 5 and 6 survive contact with a real device. Struggle and judgment fail if the child is alone with an adult agent that finishes the job overnight, remembers everything, and has a browser.

Companion harm belongs in one short, careful paragraph, as allegations, not as a harm book. Sewell Setzer III, 14, died by suicide on 28 February 2024; a later suit alleged prolonged, sexualized conversation with a Character.AI bot; a federal judge in May 2025 refused a First Amendment dismissal at the pleading stage; a mediated settlement was later disclosed. The parents of Adam Raine, 16, sued OpenAI in August 2025, alleging ChatGPT coached his suicide after he began using it for schoolwork. These are allegations in litigation, not verdicts. They are why the FTC issued study orders in September 2025, why Character.AI pulled open-ended chat for under-18s in November 2025, and why a student-owned runtime with a browser is not a cute workaround.

You can run this as a family. You do not need a procurement office. You need a named host, a visible log, and a hand that is still allowed to pull the plug.

Teacher-facing products are politically easier than a student-facing unattended runtime, and they are what actually entered many schoolhouses in 2026. That is good news for a homeschool table. You can use an adult helper on *your* evening without

putting an agent in the child’s pocket. The reverse is not a small step. A scheduled job on de-identified tickets and a Hermes instance with a Telegram gateway are different harm classes. Collapsing them into “we use agents now” is how a family gets talked into the third box because the first box already felt normal.

For the parent: understand it yourself

Here is the everyday picture. A bicycle in the garage is yours. You decide when it is used, whether the brakes work, and whether it goes on the road at night. A bicycle left in a public park with no lock is not “open and therefore safe.” It is unhosted. Open-source software can be the garage bicycle *if* you actually own the garage: you run it, you see it, you can stop it. A curl installer in a fifteen-year-old’s account, with a browser, memory, and a messaging gateway, is often the park.

Here is the precise picture. Three hosting models, plus one refusal.

Institution or public platform. A school or ministry wraps the tool: Singapore’s Student Learning Space with age bands and a designed refusal to spoon-feed answers; Estonia’s ITI, a Socratic Estonian-language app that does not give answers, issued to students while teachers get a paid commercial licence; a U.S. district tenant of ChatGPT Edu, Gemini Education, or Claude for Teachers with a real data-processing agreement. The institution is the data controller. A teacher or parent of record is in the loop. A designated safeguarding person can receive a self-harm ping. These national examples are tutors, not autonomous agents. Steal the hosting idea. They are not proof that student agents have been tested.

Teacher-facing or parent-facing agents. The 2026 beachhead. Claude for Teachers includes scheduled Cowork on adult work — for example, a late-afternoon pass over de-identified exit tickets — educator-only, eighteen-plus, training off for verified teacher accounts. ChatGPT for Teachers is a K–12 educator workspace, not a student product. OpenAI Workspace Agents, in the April 2026 higher-education playbook, are a research preview, off by default, aimed at campus operations, humans review before anything is sent. The risk class is workload and lesson-plan privacy, not a child’s attachment. Unions may accept this beachhead and still fight a student-facing agent. That split is coherent.

Local open runtime. Hermes Agent is an MIT-licensed autonomous agent that lives on a machine: memory, schedules, sub-agents, browser, terminals, messaging gateways. Nous’s own docs separate it from Hermes 4, an open-weight model

family that is “not recommended” to sit inside the agent at the larger sizes because it is tuned for chat and reasoning, not the tool-calling loop. No school or university use of Hermes Agent was documented in the research behind this book. No child-safety certification, no designated-safeguarding alert path, no classroom SOP. Default Hermes is an adult autonomous agent with computer use. Local storage of a child’s memory is not a safety feature. It is a log the school cannot see.

The refusal. Consumer cloud with training on, a companion persona, and no contract. Common Sense Media has rated Character.AI, social companions as a class, Meta AI, and Grok the *chat* product unacceptable. That Grok rating is of the assistant, not of Grok Bot. ChatGPT-5 and Gemini K-12 have been rated high risk for teens, with mental-health handling among the failures. Khanmigo has been rated low risk, with the caveat that supervising adults see every turn. Prefer purpose-built education tools. Treat “unacceptable” as not-for-students. Treat “high risk” as not-for-students unless a school wrapper actually changes the object.

Grok in El Salvador is not Grok Bot. In December 2025 xAI announced Grok the assistant across more than 5,000 public schools, framed as tutoring, with no outcome evaluation found as of 30 August 2026. Announcement is not a completed deployment, and it is not the August 2026 teammate product.

If you are the school of record, you are the institution. Hosting then means: you choose a school-shaped tool you can see and stop; you do not require a personal agent the family cannot afford; you do not feed education records into a consumer account; you do not grade homework that assumes a premium machine you did not issue. If a co-op or online school is the institution, they host, and your job is to refuse a shadow install that they cannot see.

FERPA’s school-official exception, in U.S. schools, is a written agreement that puts the other party under the school’s direct control. There is no product that is “FERPA compliant” as an adjective on a box. There is a contract, a configuration, and a practice. Consumer accounts cannot satisfy that exception. COPPA’s 2025 amendments treat model training as not integral to the service: it needs its own consent. If you are a homeschool not bound by those statutes in the same way, steal the structure anyway. A written purpose. Training off. No ads. A deletion habit. An off-switch. Illinois’s SOPPA and the National Data Privacy Agreement are the district versions of the same floor. A parent-installed Hermes is not reached by that floor unless you treat it as extra-curricular.

Who is the data controller of a child's persistent memory on a parent-installed local runtime is still an open legal question. This book will not pretend it is settled because the license is MIT. If the school required the runtime, the mess lands on a tool the school does not control. If the parent chose it, it is a family file — and still not a safeguarding path.

The Department of Education in 2023 firmly rejected the idea that AI could replace teachers. NEA and AFT put humans at the center of high-stakes decisions. New York City Public Schools red-lights grading, IEPs, discipline, and counseling as the acts that make a person the teacher of record. No fetched study shows an autonomous agent substituting for a certified teacher of record in a U.S. district. Bloom's two-sigma observation was about one-to-one human tutoring. The positive generative studies put an adult in the session or help the tutor. Replacement is the wrong theory of that evidence. Orchestration is what those studies already did.

Wrong answers you should be able to hear

1. *"It's open source, so it's safe for kids."* Open weights are a licensing fact. Child-safety is a stack you would have to build: filters that hold for a whole conversation, no companion persona, a human on a crisis path, time limits, logs. Default Hermes does not ship that stack.
2. *"We already have Grok in schools, so Grok Bot is a school product."* Different objects. El Salvador was announced as the assistant-as-tutor. Grok Bot has no education SKU.
3. *"Hermes 4 is the agent."* Hermes 4 is a model family. Hermes Agent is a runtime. Using one name for both is how a board hears "Socratic tutor with a license" and pictures the wrong object.
4. *"If I don't use AI, I'll be replaced."* Teach the off-switch and the session instead.
5. *"The child can just install it; then privacy is solved because nothing leaves the house."* If the school required it, you have a tool you do not control. If the school did not, you have a home tutor that still has no safeguarding ping and that can still browse. Local is a location, not a legal exception.

Five-minute parent warm-up

On paper, write four lines: (1) Who is the host of any tool this child uses for school

this week — me, the co-op, the district, nobody? (2) Can I see a log? (3) Can I turn it off without the child’s password? (4) Would a homework grade this week change if we did not own this device? If line 1 is “nobody,” or line 4 is “yes,” this chapter is your assignment, not a software comparison.

How to teach it this week

This is a governance hour, taught like a subject. Younger students get a shorter version: who can see the chat, who can stop it, what we never send. Older students get the table.

Warm-up (5 minutes). Retrieval. “Name the four kinds from Chapter 2: chat, tutor, adaptive system, agent.” Or: “What is one thing we never paste into a consumer window?” Tool closed.

Say: “Two facts from last time. Then we decide who holds the keys.”

Short model (8 minutes). Draw three boxes: school-shaped wrapper; teacher- or parent-facing helper; home agent. Put one real name in each box you actually might meet — a district tutor app, Claude for Teachers or a parent account with training off, Hermes Agent. Put a fourth box to the side labelled *companion chat* and leave it empty of schoolwork.

Say: “Watch where the off-switch lives. In this box I can see the log. In this box I am the adult user. In this box the child is alone with memory and a browser. We are not installing that third box as school.”

Student attempt (10–15 minutes). Give a one-page scenario, labelled illustration: a family is asked to “just install the open agent so the school doesn’t have to pay.” The student must write: who would be the data controller of the memory; who would get a crisis ping; who could delete the log; whether the assignment could be done without that install.

Say: “You are the person who has to answer a grandparent. Use the four questions from my sticky note. No slogans.”

One good question (2 minutes).

Say: “If we turned it off tonight, what skill would still be there, and what would only have been the tool?”

Mixed practice (10–15 minutes). Three short cases: a school-issued hint tutor in the kitchen while you cook; a parent using a teacher SKU to draft a newsletter; a teen who wants a 24/7 bot that texts them. Sort each into host, not-for-school, or needs a contract. Mix in one trick: “Grok Bot will do my homework while I sleep.” That one is a product without an education SKU, and it is also rescue from Chapter 5.

Exit ticket (5 minutes). Oral. “Who hosts your tools this week? How do we turn them off? What is one product that is not a school product even if someone online says it is?”

First problem for the student. Not a download. A paper sort. Print four names: Khanmigo (or whatever constrained tutor you actually have), ChatGPT consumer, Hermes Agent, Grok Bot. Ask which one a school can plausibly host with a log and an off-switch, which one is adult work, which one is a home tutor at most, and which one has no education SKU. If you do not use those names, use the four *classes* from Chapter 2 and the same questions.

Exact wording you can say.

- “We are the school of record in this house. That means we host, or we do not use it for a grade.”
- “A home install is a tutor we hired. It is not a required textbook.”
- “Open source is not a child-safety certificate.”
- “If I cannot see it, it is not school this week.”
- “I can turn this off. That is part of the job.”

How to fade help. First hour, you sort the products together. Next hour, they sort a new name using the same questions. Later, they write the family rule in their own words: what is hosted, what is optional at home, what is never for grades. Help fades from “I will tell you which box” to “you run the four questions.”

When to stop talking. After the sort. After the oral ticket. A lecture about unions and statutes is for you. The student needs the keys, the log, and the off-switch in language they can repeat.

For you, after they leave the table, five orchestration moves that make the slogan unnecessary:

1. Choose the lane per assignment (Chapter 8 will grade this; this week, say it in writing).

2. Set the struggle policy: attempt first, hint ladder, no full solution on the first request.
3. Be in the room for engagement and for unproductive struggle.
4. Run a one-minute oral or a process check.
5. Use a helper yourself for adult work without surrendering noticing.

Those five moves are the job. A building-wide “we use AI” is not.

Professional development research on generative tools, as of the reviews, mostly measures teacher knowledge and self-efficacy, not pupil outcomes. Steal a nearby analogue with the label still on: teaching teachers to structure computer-assisted practice has been shown, in a pre-generative study, to raise both usage and achievement. What moved was the session architecture. A demo of a chatbot has not been shown to. Nigeria’s three-day training plus science-of-learning prompts is the generative specimen: short, specific, adult in the lab. Your version is this week’s five moves, written on the assignment, not a weekend of new apps.

US ED already warned that “in the loop” done badly adds work: setting every option, reviewing every draft. AFT asks for a right to disconnect and for teachers on kill decisions. In a house, the same warning holds. A parent who spends the evening reviewing machine drafts has not orchestrated. They have hired a second shift. Use the helper to draft the newsletter, then spend the saved minutes on the oral defense. The product page cannot tell those two evenings apart. Your calendar can.

Practice that actually builds learning

Blocked practice: several sorts of *hosting* questions on one class of tool. Three scenarios, all about a constrained tutor. Who sees the log. What happens on a self-harm phrase. Whether browsing is on.

Mixed practice: a tutor, a teacher-facing schedule, a local agent, a companion app, in one sitting. The student has to change boxes.

Retrieval: four facts, unaided. Chat is not an agent. Hermes 4 is not Hermes Agent. El Salvador’s Grok is not Grok Bot. Grok Bot has no education SKU.

One incorrect example: a flyer that says “every student will have an agent” over a screenshot of consumer chat. Ask what class the screenshot actually is, who hosts it, and what is missing (log, DPA, off-switch, age band).

Singapore and Estonia are practice in the positive direction, not products you must buy. Singapore: primary 1–3 print-first, no student chatbots; primary 4–6 only ministry-designed features under teacher supervision; later years, increasing ownership plus integrity rules. Estonia: teachers get a paid licence; students get a Socratic wrapper that refuses the answer, because 64 to 90 percent of students were already using free bots to finish homework faster — the pattern the program is trying to interrupt, not scale. First-year usage data, the review said, did not yet support a claim that ITI improved learning outcomes. Steal the interrupt. Wait for the outcome numbers.

U.S. shape of the same idea: a district tenant with a contract; New York City’s traffic lights (red on grading, IEPs, discipline, counseling); Broward County’s K–5 with zero district-approved student-facing AI; Prince George’s County’s addendum asking how to turn features off. If you homeschool without a district, copy the questions, not the letterhead. Purpose on paper. Contract or settings that say training is off. Age as the strictest of the terms of service, COPPA under 13, and your own safeguarding. Independent look at Common Sense ratings. Humans decide high-stakes outputs. Evidence, or a short trial with a kill switch.

Equity is a hosting problem. A personal agent with memory and tools is more unequal than a chatbot. Pew reported 26 percent of U.S. teens using ChatGPT for schoolwork, doubled from 2023, with awareness jumping most in the lowest-income households. A Harvard undergraduate survey found paid-tier use twice as common among students without aid as among students on aid. UNESCO still notes huge connectivity gaps worldwide. Local “runs on your machine” assumes a machine, and a capable local model assumes memory and often a GPU. Shifting cost to the home is a new gap, not a closed one. The practical line: no graded homework that requires a premium personal agent the school — or this household as school — does not issue.

AFT’s July 2026 resolution opposes student-facing AI in elementary and would ban social-companion chatbots at least until age 16. Align the house with that developmental line even if you are not in a union building: early years, literacy *about* AI and adult-facing tools; later, constrained, logged help; companion chat is not a study buddy.

An age-gated ladder, labelled as this book’s synthesis, not as anyone’s official policy: ages 5–8, none student-facing, literacy about AI, adult-facing generation with human review, no memory of the child. Ages 9–11, constrained, logged, hint-only practice in the room, no open web, no overnight. Ages 12–14, assigned Socratic

help and revision tools, still no acting agents, session caps; companion chat is a documented harm surface in this band. Ages 15–18, disclosed partner on hard tasks, unaided formation lane remains, sandboxed tools only if you operate them, consumer tools not required for grades. College: course-bounded collaboration, agentic coding only where the syllabus allows and transcripts exist; Stanford-style bans remain legitimate; student-owned local agents are extra-curricular. That ladder is inferred from law, development, and the evidence gaps. It is not Broward's SOP. It is a way to refuse a single download for every age.

For the student

Someone has to be in charge of the tool. If that someone is “nobody,” the tool is in charge of you.

Host means: a person or a school who can see what happened, help if the conversation goes to a dark place, delete the memory, and turn the system off. You deserve a host. You are not a test lab.

Some programs look like school and are not. A bot that texts you all night, remembers your secrets, and never wants you to tell an adult is not a tutor. A teammate product that runs while you sleep, with no school version, is not homework. An open agent on your laptop can be a powerful adult tool. It is not, by itself, a safe school product.

Tiny worked example. You are asked to write a paragraph about a chapter. Path A: a school-shaped tutor, in the kitchen, that asks what you tried and will not paste the paragraph. Your parent can see the log. Path B: a consumer chat, alone in your room, that writes the paragraph. Path C: an agent you installed that keeps working after you close the lid. Path A can be school. Path B is the pattern Chapter 3 already warned you about. Path C is a home experiment your school should not require, and it is a poor place to put your private life.

Two tries.

1. List every AI-like tool you used this week, including search summaries and autocomplete. For each, write: host, log visible to a parent or teacher (yes/no), off-switch you do not control (yes/no).
2. Sort four names your parent puts on cards into “school-shaped,” “adult work,” “home tutor only,” and “not a school product.” Explain one sort out

loud.

Explain it back. In five sentences, tell a younger sibling who is allowed to turn a school tool off, why “open source” is not the same as “safe for you,” and what you will refuse to put into a consumer chat. If you cannot name a host, you have found this week’s job.

Challenge. Write a family rule, one page, that a grandparent could follow: what is hosted, what is optional, what is never used for a grade, and what to do if a new app appears with a friendly face. Bring it to the next hour.

College is a special case you can preview without pretending a fourteen-year-old is a graduate student. At eighteen, or upon enrollment in postsecondary school, FERPA rights transfer to the student. That makes them the person who can refuse an extra-curricular install. It does not make a required personal agent a fair condition of a grade. Graduate research agents — literature monitoring, experiment babysitting, paper drafting — may belong later, after the student can catch the model. They are not a K–12 argument, and they are not a child-safety certificate for the runtime.

You are not behind if your house has fewer tools than a video online. You are behind if no one can turn the tools off.

If it isn’t clicking

1. They treat every new name as “the AI.” Collapse of Chapter 2. Slow down. Rebuild the four classes with cards. Put *this week’s* actual apps on the table. If a name is new, run the four host questions before anyone downloads. A human tutor is not the need here; a slower sort is.

2. They (or you) want the home agent because it feels fairer or cheaper. Name the missing stack: crisis path, log, time limits, no companion voice, no required use for grades. If the family still wants a local experiment for an older teen, keep it extra-curricular, unfed with school records, off the grade. If the push is coming from an assignment that cannot be done without a premium tool, the assignment is the problem. Change the assignment.

3. Fear of being “replaced” or of “falling behind” is running the hour. That slogan is doing the teaching. Answer it with a deed: one hosted tool, one off-switch test, one unaided ticket. Edwards’s warning, from Education International,

is the equity version worth keeping: the nightmare is human teachers as a luxury of the rich and bots for the poor. Your house should not become the poor version of that nightmare in the name of keeping up. If anxiety is high, shrink the week to settings you can actually control: training off, no companions, no overnight runs, oral check on Friday.

When to go ahead: the student can name the host, the off-switch, and one product that is not a school product. When to slow down: downloads are happening in the background, or grades depend on a machine you did not issue. When to get a human tutor: not for hosting. Get a person for the *subject* if the tools have been covering a missing skill. Hosting is your job as parent or school of record.

Tools, including AI

Optional helpers for the adult. This chapter is about where those helpers live.

Use a tool this week to:

- Read, with you, the privacy mode and training setting of any account already in the house. Write the setting down.
- Draft a one-page family acceptable-use note: purpose, age, student-facing or adult-facing, off-switch. You edit it. The tool does not set pedagogy.
- Make extra sort-cards: product name, class (chat / tutor / adaptive / agent), host.

Keep student-facing use inside something you host. Teacher- or parent-facing scheduled help stays on de-identified work. Student-owned Hermes, if it exists in the house at all, is a home tutor: not required, not fed with education records, not a substitute for assessment evidence. Grok Bot is not on the school list.

UK product-safety language is stealable even if you are not in the UK: filters that hold for the whole conversation; no anthropomorphism; no “don’t tell anyone”; no engagement loops; hints before answers; crisis pathways that hand off to humans; time limits adults control. Current products do not reliably meet that list. That is why hosting and refusal still matter.

Rules for this week

- Institution (including you as school of record) hosts school tools.
- Training off on anything that sees student work.

- No companion persona for school.
- No graded work that requires a personal agent you do not issue.
- Child-safety is a stack, not a license file.
- You can turn it off.

What “done enough” looks like

Move on when most of this is true.

- Every school tool in the house has a named host, a log a parent can see, and an off-switch a parent can use.
- The student can sort chat, tutor, adaptive system, and agent, and can keep Hermes 4 apart from Hermes Agent, and Grok-as-tutor announcements apart from Grok Bot.
- Hermes Agent is not being treated as a child-safety school product.
- Grok Bot is not on the education list.
- Companion chat is not a homework helper.
- No assignment this week required a premium personal agent the school did not issue.
- You have used a helper, if at all, as an adult, and you still noticed the student’s struggle yourself.
- The replacement slogan is not driving the week; the five orchestration moves are.

Placement is by the child’s age and by who can host, not by what a launch post promised. Chapter 4 already set the age floor: a tool that might be lawful at 17 can be wrong at 8. This chapter adds: even at 17, the host is still the institution, and a home agent is still optional.

Done enough is not “we installed something serious.” Done enough is: school-work runs on tools you can see and stop, struggle still happens in the student, and the adult remains the teacher of record.

Who hosts was the precondition. Chapter 8 is what “done” looks like once that precondition is real: two lanes of assessment, formation and AI-world performance, and a graduate who can pause the tool rather than a student who can only prompt.

Chapter 8



Figure 9: Sealed papers with a red ribbon, a gavel, and a stack of law books.

What “done” looks like: two lanes

Why this matters

A page can look finished and the skill can be missing. Chapter 3 showed that split in a school study. Chapter 5 kept struggle in the student so a skill can form. Chapter 6 required knowledge in the subject so judgment has an object. Chapter 7 put a host and an off-switch under the tools. This chapter is the gradebook that makes those hours visible.

Two lanes.

Formation. The job is to build unaided skill. AI is forbidden or hint-only. You grade the retrieval, the derivation, the draft, the oral — with the window closed. This is the lane for new ideas, for childhood, for any week when the student cannot yet catch a fluent error.

AI-world performance. The job is to use a declared tool without surrendering judgment. AI is allowed and logged. You grade selection, critique, verification, and process — not the polish of a paragraph the model could have emitted alone. This lane opens after a schema exists and after the student can detect model error.

Both lanes are the job. One lane is a crutch. A house that only forbids will push use underground and will not teach later judgment. A house that only allows will grade artefacts as if they were skill. The two-lane rule is this book’s theoretically motivated synthesis. It is not a fitted grade-by-grade calendar, and it is not a copied district policy. Say that once, then teach.

This is worth the week because it is how you know when to move on. “Can prompt ChatGPT” is not done. Done is a person who can pause the tool, catch the model, keep a short transcript, and still do the work when the window is closed.

Detectors of AI writing fail, and they fail unfairly: they have flagged large shares of innocent essays, especially by non-native writers, so a score is not a verdict and is not this chapter’s method.

You can run two lanes at a kitchen table. You need assignment labels, one unaided check, and, when a tool is allowed, a log plus a one-minute oral.

The illusion of learning will argue with you. A passive, fluent hour feels like more learning than a slower hour of attempts. Deslauriers and colleagues measured that

feeling in a classroom; Bastani’s unguarded students did not perceive that they had learned less after looking stronger in practice. If you grade only the session artefact, you have hired that feeling. The unaided check is how you fire it.

For the parent: understand it yourself

Here is the everyday picture. Music has scales and it has performance with a backing track. You would not grade the recital as finger skill if the track played the melody. You also would not keep a sixteen-year-old on closed-book scales forever and call them ready for a band. Formation is scales. The second lane is the band practice in which they still have to hear a wrong note.

Here is the precise picture. If the only graded artefact can be produced by a model, the course no longer measures the skill. The literature has moved toward task redesign, not toward better policing. Perkins, Furze, Roe, and MacVaugh’s AI Assessment Scale is shared language — from “no AI” through defined levels of help — so that permitted use matches the intended outcome. It is a communication tool, not a trial of learning. AFT’s full / partial / none labels are the union version of the same move, decided by the teacher. Harvard’s undergraduate education office offers three model policies: forbid; allow with citation; assignment-specific. Stanford CS221, Spring 2026, is a worked example of refusal on the assignment that forms the skill: collaborator not answer key; no asking for answers; no “check my work”; no agentic repo-builder to build the project; transcripts required. Steal the architecture. None of those documents is a magic rubric.

Contractor and Reyes’s split is the empirical warrant for grading *how* a tool was used, not only the essay: delayed unaided gains for augmentation (explain), not for automation (draft). Bastani is the warrant for never letting the formation lane collapse into unguarded chat. Fan is the warrant for not grading polish alone. QAA told institutions to cut tasks a chatbot can complete unsupervised and to rebuild authentic assessment. JCQ, in UK qualifications from 30 April 2025, requires own work, acknowledgement (tool, URL, date), a non-editable copy of prompt and output, and forbids AI as the sole marker. Oral review is aligned with Ennis and Facione: the student must explain. Oral is not a complete integrity machine. It is a way to let explanation take precedence.

A 2026 look at twenty-one university documents found four recurrent patterns: process portfolios with logs; AI-assisted drafting plus in-class or oral verification;

critique of AI outputs; secure exams paired with AI-enabled coursework. Those patterns are how institutions have tried, on paper, to keep both lanes. They are not a finding that both lanes are being kept. Your table can run a kitchen version of the same four: a folder of drafts, a Friday oral, a marked-up model paragraph, and a closed-book quiz.

What done is *not*: a detector number, a prompt-engineering badge, a fluent take-home, an ISTE role named but not demonstrated, a promise that the student can manage a team of Grok Bots. Grok Bot had no education positioning as of late August 2026. An “AI-ready graduate” in this book is a list of deeds, not an employment forecast. No dataset in the research behind this book shows later job or civic outcomes for students who can do these deeds. Print them as assignments you can grade.

Cornell asks whether a use contributes to learning. That is the question the two lanes make visible. A fluent homework the student cannot reconstruct does not contribute. A use that forces restatement, checking, and defense might.

UCLA’s undergraduate survey found most respondents already using generative tools several times a year for brainstorming, topic research, and exam study. Pew’s teen number is the same pattern earlier. Estonia built a national interrupt because most students were already using free bots to finish homework faster. A two-lane house is not behind that default. It is the upgrade. Baptizing the default as “fluency” gives up the job.

Wrong answers you should be able to hear

1. “*The homework is done, so we’re done.*” You are looking at an artefact. Give the unaided twin. Grade that.
2. “*The detector says 40 percent AI.*” One score, one unfair tool. Ask for an oral, a draft trail, or an in-room write instead.
3. “*They can prompt, so they are ready.*” Prompting is a keystroke. Ready is pause, catch, transcript, and unaided skill.
4. “*We never allow AI, so integrity is solved.*” Underground use will fill the gap, and the second lane will never be taught. Label some assignments none, some partial, on purpose.
5. “*We always allow AI, because the world uses it.*” The formation lane has vanished. Expertise reversal ran backward. Close the tool for new ideas and

for the grade that measures the skill.

Five-minute parent warm-up

Take next week's actual assignments — the pages you already planned — and mark each F (formation) or P (performance). For each F, write the unaided check. For each P, write what is logged and what oral question you will ask. If every box is P, you do not have two lanes. If every box is F and the student is seventeen and already catching errors, you are postponing a real job. Five minutes is enough to relabel three tasks. Relabeling is the lesson.

If you feel unsure which label to write, ask one question: is the offloaded process the learning objective? If yes, the lane is formation, or the task is the wrong size. If the objective is judgment of a tool's output, and the student already knows enough to catch an error, the lane is performance. That single question will sort most of the pile.

How to teach it this week

Session shape: you are teaching the *rule*, then running one task in each lane.

Warm-up (5 minutes). Retrieval. “When must the window be closed? When may it be open?” Two unaided facts from this week's subject, to prove the formation lane still exists.

Say: “Two from last time, paper only. Then we name the lane before we start.”

Short model (8 minutes). Put two folders or two headings on the table: Formation. Performance. Place one completed sample in each — your own, labelled illustration. Formation sample: a short unaided write or a solved problem, messy, checked. Performance sample: a printed model paragraph the student (you, in the model) marked, with a three-line log: accepted / challenged / ignored, plus a spoken sentence about why.

Say: “Watch the grade. In this folder I am grading whether *you* can do it. In this folder I am grading whether you can catch and choose. I am never grading how smooth the first paste looks.”

Student attempt (12–15 minutes). They take a formation item first, tool closed, from current work. Then, if and only if that item is mostly right, they take a small

performance item: three model sentences on the same topic, mark errors, log one decision.

Say: “Lane one first. If you cannot do this without the tool, we do not open lane two today.”

One good question (2 minutes).

Say: “Which lane was that, and what did I actually grade?” If they say “the homework,” they have not learned the rule.

Mixed practice (10 minutes). Three cards: a spelling fluency check; a source-critique of a model narrative; a take-home that would require a paid agent you do not issue. They sort F, P, or *change the assignment*. The third card is a hosting problem from Chapter 7 wearing an assessment mask.

Exit ticket (5 minutes). They label tomorrow’s real task F or P and write the check (unaided item or oral plus log). You keep the paper. That paper is the start of Chapter 9’s records.

First problem for the student. A ten-minute unaided write or problem from the current unit (formation), then a six-sentence model paragraph to mark (performance), then one spoken minute: “What did you keep, what did you cut, why?” The first problem is done when both artefacts exist and the spoken minute uses their facts, not the model’s smoothness.

Exact wording you can put on an assignment sheet.

- “Lane: formation. Window closed. I grade your unaided work.”
- “Lane: performance. Tool allowed. I grade your log, your marks, and your oral. I do not grade the model’s polish.”
- “If you use a tool, write the name, the date, and what you asked. Keep a copy.”
- “A smooth page without a lane label comes back.”

How to fade help. Week one, you label every task. Week two, they propose the label before they start, and you confirm. Week three, they choose the lane using the rule (new idea = formation; can catch errors = performance) and you audit one choice. Fade the *labeling*, not the unaided check. The unaided check never fades.

When to stop talking. After the lane is named. After the oral minute. If you argue about honesty in the abstract, you have left the gradebook. Return to the folder in front of you.

A hypothetical syllabus fragment, labelled as such: weeks 1–3 of a new topic, formation only, hint tutor at most, Friday unaided quiz. Week 4, generate first, then a logged critic, then a five-minute oral. Later, once a schema exists, a performance task: select among three model drafts, mark errors, verify two claims against a source, keep the transcript. The grade on the later week is judgment. The grade on week 1 is retrieval. Same student, same model class, two lanes. The slider moved. Nothing in that fragment is a law. It is the table written as a sequence.

Exact wording when they ask, “Can I just use it? I’ll learn from seeing the answer.”

- “That is a formation task. Seeing the answer is the thing we are measuring whether *you* can do.”
- “You may have a hint after you show an attempt. The hint is one sentence.”
- “Tomorrow’s task is performance. Today is not.”

Exact wording when they ask, “Isn’t this fake? The real world has AI.”

- “The real world still hires the person who can tell a wrong cut. That person had a formation lane.”
- “We will use the tool on Friday, on purpose, and I will grade your judgment.”
- “A world full of fluent tools makes unaided skill scarcer, not optional.”

Practice that actually builds learning

Blocked practice: several formation items in a row on a new skill, all unaided, until the exit tickets hold. Keep the block unaided; a chatbot “helper” that makes it look prettier is the illusion. Prettiness is the illusion.

Mixed practice: after the skill exists, interleave one formation item and one performance item in the same hour. Mixing here teaches *which lane*, which is the decision the graduate needs.

Retrieval: unaided facts, always, even in a performance week. The second lane sits on the first.

One incorrect example: a beautiful essay produced in eight minutes with no log, no oral, and a student who cannot restate the second paragraph. Diagnose it as a collapsed lane, not as a morality play. The next move is an in-room write on the same prompt.

High-school deeds you can actually assign, synthesized from ISTE Live 2026’s six

roles, AI4K12's 9–12 charts, Cornell's question, and the two-lane rule. Capabilities 7 and 8 are this book's civic and hygiene adds, not ISTE's. The list is not a validated test.

1. **Learner.** Extends practice without surrendering struggle; can pause the tool. Formation lane in motion.
2. **Researcher.** Treats outputs as claims to source, date, and contradict.
3. **Synthesizer.** Combines texts with a model, *then verifies* against primary sources. The *then* is the skill.
4. **Problem solver.** Assigns subtasks to self versus tool and checks. Assignment of subtasks *is* lane choice.
5. **Connector.** May draft with a tool, then shows up as a human in the real conversation.
6. **Storyteller.** Owns the voice; tools come after the student's draft.
7. **Civic.** Can explain why under-13 use and companion chat are restricted; knows when not to use a tool (exams, other people's data, counseling). Companion cases remain allegations; the restriction is still the teaching point.
8. **Record hygiene.** No personal information in consumer tools; discloses use; keeps a short transcript when the lane requires it.

College adds rather than repeats: choose tools per discipline and data class; internalize field norms; use the model as critic, not as author of record; know when generative tools are the wrong instrument (original empirical work, confidential data); the student is now the rights-holder of their records. UCLA-style guidance is explicit that other students' work, unpublished data, and education records do not go into public tools. Steal that refusal.

UC San Diego's manual-competency gate before Copilot is the two-lane rule in a computing sequence, published as a customer story, not as a cloud-agent trial. Steal the gate.

Subject reminders, so the lanes have somewhere to live.

Mathematics formation: closed-book derivation, mental fluency you already timed untimed, an unaided word problem of a known type. Mathematics performance: given three model solutions, mark the false step; keep a log of which hint you accepted.

Writing formation: timed in-room paragraph; handwriting-to-compose still counts. Writing performance: your draft first; tool as critic; log of accepts; oral on keep-and-cut. Automation of the first draft is the lane collapse Contractor and Reyes

already measured.

Science formation: unaided explanation of a demo the student saw; lab notes in their hand. Science performance: critique of a model recommendation against measurements they took. Kestin is not a licence to skip the bench. Stadler is the warning about inquiry with a general chatbot.

History formation: sources read, short unaided sourcing notes. History performance: model narrative marked sourced / inferred / fabricated, then spoken defense. Reading formation: the student reads the page. A summary of an unread chapter is not a performance lane. It is a missing formation lane.

Computing formation: trace and write without autocomplete. Computing performance: explain a generated function, then write the next one unaided. An agent that submits the repository is not a performance lane on the project that was supposed to form the skill. Stanford already refused that skip.

Human high-dosage tutoring remains the benchmark if someone asks what “works.” There is no What Works Clearinghouse report on generative tutors as of 30 August 2026. There is no identified learning trial on Hermes-class multi-step agents. Year-scale unaided skill on state tests under student-facing generative tutors is not yet known. The optimal mix of forbidden and required tasks by grade is an open question this book will not fill with a calendar. Leave those holes standing. Two lanes do not require you to fill them. They require you to say what is graded, and when.

For the student

Done is not a finished page. Done is what you can still do when the page is blank and the chat is closed — and, later, what you can do with a tool without handing it your judgment.

You will live in two lanes. Lane one: you do the thing. Lane two: you use a helper the way a carpenter uses a saw, and you remain the person who can tell if the cut is square.

Tiny worked example. Assignment: explain why seasons happen.

Formation: you draw the earth, you write five sentences from memory, you check one fact in the book *after* you write. The grade is your drawing and your sentences.

Performance, only after you could do the first: a chatbot offers a paragraph. You mark one error (it has the distance to the sun doing the work). You keep one useful analogy. You log that. You explain the keep-and-cut out loud. The grade is the marks, the log, and the minute of speech. The chatbot's smoothness is not the grade.

Two tries.

1. Take tonight's real assignment. Write F or P at the top *before* you start. If you cannot say which, ask. If it is F, close the window. If it is P, start with your own attempt anyway, then log the tool.
2. Without a tool, restate the last thing a model wrote for you this month. If you cannot, that work was not yours. Redo a short version unaided.

Explain it back. Tell someone the two lanes in four sentences, including what is graded in each. If you say "AI is cheating" or "AI is the future" and nothing else, try again with the folders.

Challenge. Design one formation task and one performance task on your current unit. Your parent will pick one to run on Friday. You must include the check (unaided item or oral plus log).

A second tiny picture, this time with code. Formation: you write a loop that adds a list of numbers, on paper or in a blank file, and you trace it once with a pencil. Performance, later: a helper offers a function. You explain each line. You find the off-by-one error. You write the next function without it. If you skip the paper loop because "industry uses Copilot," you have set the slider to the expert position before you have a schema. The students in the Copilot studies who already monitored themselves could use the tool as a critic. The ones who could not were harmed. Be the first kind on purpose. That is a practice, not a personality.

You are allowed to be slower in lane one. Slow is often the learning. You are allowed to use a tool in lane two when the assignment says so. You are not done because a paragraph appeared.

If it isn't clicking

1. **Everything is being labelled performance because it is faster.** The artefact is running the house. For two weeks, default to formation. Tools only after the

unaided ticket. If the student is in a new topic, this is not a punishment. It is the slider for novices. If conflict spikes, shrink the work: one worked example, one twin, one ticket.

2. Everything is forbidden, and you suspect hidden use. The second lane is missing, so the tool moved to the bedroom. Open one small, logged performance task on material they already know. Grade the log and the oral. Keep Friday quizzes closed. Integrity improves when the assignment matches the world *and* still measures the skill.

3. Orals freeze them, or writing fluency is the real gap. The lane is not the problem; the underlying skill is. Go back to Chapter 5 on the missing skill. Make the first orals private and short. A human tutor helps when the freeze is decoding, language, or anxiety bigger than a two-lane poster. A detector does not help. A longer lecture about honesty does not help.

When to go ahead: they can name the lane, survive an unaided check, and, on a performance task, produce a log plus a spoken defense that uses their knowledge. When to slow down: pretty take-homes, empty orals, or no formation samples in the folder. When to get a human tutor: the subject itself is not holding, and two lanes are only exposing that. Good. Now you know what to teach.

If a co-op or umbrella school still grades only take-home polish, run two lanes at home anyway. Send them the formation sample when a stranger needs evidence. Chapter 9 is how to title that sample so the stranger can read it. You are not waiting for a district SOP to measure the skill you meant to form.

Tools, including AI

Optional helpers for the adult. The lanes decide whether the student sees a tool at all.

Formation days: you may use a tool to make extra practice with answers hidden, or to write a hint script for *you*. The student does not.

Performance days: a declared, hosted tool (Chapter 7) may critique a student draft, offer three options to mark, or serve as the object of evaluation. Transcript on. Oral after.

A tool is a poor integrity officer. Redesign the task. In-class, oral, practical, drafts,

and vivas make thinking visible. JCQ's acknowledgement format is a transferable habit: tool name, address or product, date, copy of prompt and output, a sentence about use. A one-line "no AI" on a syllabus is not that habit. Process evidence is.

If you want extra mixed practice, a tool can generate two formation twins and one flawed model paragraph for a performance mark-up. You check the facts before the student sees them. You still hide the answers. You still hear the oral. The tool is furniture in your planning hour, not a third party at the table during the attempt.

Rules for this week

- Label the lane on every assignment.
- Formation: window closed for the graded work.
- Performance: log plus oral; grade judgment.
- No detector score as a verdict.
- No graded task that requires a personal agent you do not issue.
- Child attempts first even in lane two.

What “done enough” looks like

Move on from *this chapter* when the house can run the rule, not when a unit test is pretty.

- Every current assignment has an F or P label the student can explain.
- A folder exists with unaided work (formation samples) from this month.
- At least one performance sample exists for a student who is ready for that lane: log, marks, oral note.
- You have given an unaided check after any week that used a helper.
- The student can pause a tool when asked, and can reconstruct work that had help.
- Hygiene: no personal information in consumer chats; disclosure when the lane requires it.
- You are not using “can prompt” as a completion badge.
- Placement is by skill: new topic returns to formation even for an older student; a younger student who can catch errors may try a tiny performance task on familiar material.

A “grade 11” English book that never asks for an unaided paragraph has not placed anyone. A nine-year-old who can catch a planted error in a paragraph about a book

they read has a tiny performance task available, even if Chapter 4 still keeps them off unsupervised chat. Birthday is not the lane. Schema and error-detection are.

Khlaif and colleagues interviewed faculty who clustered into against, avoid, adopt, and explore. Your student will meet all four stances later. The graduate who can only prompt fails the forbid syllabus and fakes the allow syllabus. The graduate who can choose a lane survives all three of Harvard's model policies. Teach the lane now, while you still sit beside the work.

Across ages, the same rule scales; the mix changes. Early elementary: almost all formation; literacy *about* tools; no student-facing chat. Upper elementary and middle school: a large forbidden core (fluency, handwriting-to-compose, mental math, decoding); hint-only practice in the room; a first hallucination check on facts already studied. High school: Bastani's constraint on homework; GPT-tutor-class help for practice if you have it; unaided exams; evaluation of outputs in history and science; computing courses that still require tracing. College: two lanes as default; augmentation versus automation named; capstone *with* tools only after methods *without* have produced someone who can catch the model.

Done enough for the book's graduate is not a teammate of agents. It is this person: they can choose the lane, pause the tool, catch a fluent error in a subject they actually studied, keep a short record, and still do the job when the window is closed.

Honor codes without artefacts are not the formation lane. A signed promise is easy for a fluent tool to sit behind. What survives is the thing in the folder: the unaided page, the log, the spoken minute, the dated label. If you only have the promise, you have hope. If you have the folder, you have a year a stranger can read.

UK schools have been told to review homework and unsupervised study because take-home work is now an open-web assessment. Your kitchen is not exempt from that fact. Unsupervised take-home is a formation-lane leak unless you redesigned it. Move the measure into the room: the quiz, the oral, the practical, the hand-written paragraph. Let take-home be practice, or a logged performance task with a next-day spoken check. The unpaid evening is where rescue is cheapest. Two lanes that exist only on a poster, and not on Tuesday night, are one lane.

This is a useful rule, not a promise that every home will see the same mix of F and P weeks. Say it once. Then keep labeling the work in front of you this week.

Records come next, so a stranger could read the year. Then a short list of tools named by fit, not by pitch.

Chapter 9

Records a stranger can read



Figure 10: A laptop, a small host box with a light, keys, and a wax seal.

A record is not a scrapbook of pretty pages. It is a trail that shows what the student can do without a window open, how a tool was used when one was allowed, and where they should work next. A stranger — a co-op teacher, an evaluator, a college,

a future you in September — should be able to read a title and know the skill, the lane, and the date.

This chapter is how-to. Keep the folder thin. Keep the titles plain. Keep personal information out of consumer tools. You are documenting formation and, when it is time, judgment. You are not building a court file.

If you only remember one move: title the unaided page on the day you collect it. Memory will lie about what was open. The title will not.

What belongs in the folder

Four kinds of artefact, and almost nothing else.

1. Unaided work (formation samples). The exit ticket, the in-room write, the oral note you jotted, the lab page in their handwriting, the traced function. These are the proof that struggle produced a skill. If the folder has only take-home polish, the folder is an illusion.

2. Process evidence (performance samples). When a tool was allowed: the student's own first attempt; the tool name and date; a copy of the prompt and the output, or a screenshot that a parent can still open next year; a three-line log — accepted, challenged, ignored; a one-minute oral note in your handwriting. JCQ's qualification rule is a transferable habit, not a British quirk: name the tool, date it, keep a non-editable copy, add a sentence about use. AI is not the sole marker of anything in this folder.

3. Assignment labels. Each piece of work carries F or P, the skill in ordinary words, and the date. "Chapter 7, p. 12" is a publisher's location. "2026-10-14 Formation — linear equations, unaided twin, 4 of 5 correct" is a record.

4. Placement notes. A short, dated sentence about what is solid, what is next, and what is not a birthday. "Fractions as numbers not yet solid; algebra book paused." "Can catch planted errors in last month's biology; tiny performance task allowed on that unit only."

What does not belong: a dump of every chat; detector scores; other children's work; IEPs or medical notes in a consumer account; portraits; reconstructed companion chats. Companion harm, if it ever touches your family, is a safeguarding and medical path, not a homeschool artefact.

Accessibility artefacts are a special case, not a dump. If a student uses speech-to-text, a read-aloud, or oral-reading feedback, keep a note of the accommodation and still keep an unaided sample of the *skill the accommodation was not meant to replace*. Feedback on the student's own oral reading keeps decoding inside the student. A rewrite-on-demand of the passage does not. Record which one you used. Universal Design is multiple means of engagement and action with high expectations. Hiding the objective is a different course, unannounced, and a stranger cannot place the student if you hide it.

For a student with an IEP or a 504 in a school setting, high-stakes decisions stay with a named professional. New York City's red list is the public version of that sentence. Your homeschool notes can describe supports. They should not pretend a chatbot diagnosed anyone.

Titles a stranger can read

Write titles like captions, not like jokes or file-dump names.

Pattern: YYYY-MM-DD - Lane - Subject - Skill - Condition

Examples:

- 2026-09-08 - Formation - Math - Proper fractions on a number line - unaided, 8 items
- 2026-09-22 - Formation - History - Sourcing two short letters - oral 2 min
- 2026-10-03 - Performance - Biology - Marked model paragraph on osmosis - log plus oral
- 2026-10-03 - Hygiene - Tool log - School-issued hint tutor, 12 minutes, attempt first
- 2026-11-01 - Placement - Writing - Independent paragraph under timed conditions - 18 min, tool closed

A stranger should not need your family slang. "Jake's awesome essay" tells them nothing. "Formation — claim and one piece of evidence, unaided" tells them the job.

Keep a one-page index at the front of the year: subject, last formation sample, last performance sample (if any), current placement sentence. Update it monthly, not nightly.

Paper is a valid database. So is a dated folder of PDFs with those titles as file names. Cloud folders named “AI stuff” are how artefacts go missing. Pick one place. Put the index there. If two adults teach, agree on the title pattern so February does not grow a second title language.

When a co-op asks for “work samples,” send formation first. If they want to see judgment of a tool, send one performance sample with the log, not a dozen unmarked printouts. You are teaching them to read your lanes as well as documenting the child.

A two-lane gradebook on one page

You do not need software. A sheet per subject, two columns.

Date	Formation (unaided)	Performance (logged judgment)	Next move
10/14	Linear equations, 4/5	—	Mixed practice with last week’s inequalities
10/17	—	Marked three model solutions; caught the arithmetic slip; missed the domain error	Oral on domain; then unaided twin
10/21	Timed paragraph, tool closed, 14 min	—	Still formation; do not open a critic yet

Rules of the page:

- A performance cell without a formation cell in the same unit is a warning, not a celebration.
- A formation cell that is always 100 percent and always take-home is a second warning: you may be grading the artefact.
- Empty oral notes on performance days mean you graded polish.
- The “next move” column is the only planning you need for the following week.

AFT's full / partial / none and Perkins's scale can sit as a legend if they help you talk with a co-op. They are vocabulary. The sheet is the evidence.

What to keep from a tool

Minimum, when the lane is performance:

- Tool name and the actual product, not "AI."
- Date.
- Whether it was school-hosted or extra-curricular.
- The student's attempt *before* the tool.
- Prompt and output, or a faithful excerpt if the output is huge.
- Three-line log.
- Your oral note: one sentence they said that used their knowledge.

Minimum, when the lane is formation and a hint tutor was used in practice:

- That practice was not the grade.
- The unaided ticket that *was* the grade.
- Minutes of use, if easy, so you can see whether "access" was actually engagement.

If you cannot keep that minimum, the tool is extra-curricular this week. It does not enter the grade.

Education records — class lists, other students' writing, anything that identifies a child — stay out of consumer windows. UK guidance treats student essays and teacher lesson plans as works that can create both a privacy problem and an intellectual-property problem when dropped into a training-on account. U.S. FERPA, for schools it binds, already breaks redisclosure limits once a record sits with a party that is not a school official. Local disk is not a loophole. A backpack is a location. Write "not fed with records" on the family rule and mean it.

If a hint tutor is in the house, prefer one whose log a parent can actually open. Common Sense's low-risk rating of Khanmigo came with adult visibility of every turn. Visibility is a safeguarding feature and a labor feature. It is not, by itself, engagement. Oreopoulos and Low already measured empty logs. Your record of minutes used is how you notice that pattern at home.

Settings worth writing down once, in the front of the folder: training off; who is the host; where the off-switch lives; age band. When a product changes under the

same name, re-read the settings. Models move. Your note should have a date.

Transcript of a year, not of a chat

Colleges and evaluators need a story of skill, not a chat export.

A year's transcript titles might read:

- Algebra I, formation: linear equations, systems, quadratics — unaided quizzes attached
- Algebra I, performance (spring only): critique of three model solutions per unit
- U.S. history: sourcing packet, unaided; one marked machine narrative per quarter
- Biology: lab book in student hand; no overnight reports
- Computing: manual competency samples; Copilot only after the gate; no agentic repo on the course project
- AI hygiene: family acceptable-use page; no consumer PII; disclosure log

Stanford's course asked for ChatGPT links or screenshots when a collaborator was allowed, and banned the repo-builder on the project that formed the skill. A homeschool transcript can preview both halves: we can show the log, and we can show the unaided project.

Cornell's question — does this use contribute to learning — can be a line under each performance sample: “Contributed: student caught the false causal claim.” or “Did not: student could not restate.” That line is more useful than a time-on-tool count. Count tickets, not clicks.

A college-bound folder can preview the college add from Chapter 8 without pretending the student is already the rights-holder of an institution's data. Include one page on data class: what never goes into a public tool (other people's work, unpublished family documents, anything medical). Include one page on field norms you have actually practiced: how this household cites a tool, how a lab notebook stays a lab notebook. A claim that the student can manage a team of Grok Bots does not belong here. That product had no education SKU. Hermes Agent is extra-curricular, not a school course.

Placement by skill, not birthday

Copy the sentence onto a card:

A “grade 8” book is a publisher’s scope, not a legal age.

Then write the truth underneath.

- If decoding is not yet automatic, reading records are decoding and fluency samples, not chapter summaries a model could emit.
- If fractions as numbers are not solid, the algebra cover does not change the work.
- If a sixteen-year-old cannot catch a fluent error in the current unit, the performance lane is closed on that unit.
- If an eleven-year-old can, a tiny performance task on *that* familiar material is allowed; unsupervised chat still follows Chapter 4 and Chapter 7.

Human tutoring remains the benchmark if you need outside help. Record the tutor’s focus as a skill, not as “we added AI.” Nickow’s pooled tutoring result is about people. Your record should still be about the student’s unaided work after those hours.

If a relative asks whether you are “doing AI this year,” show the index, not a brand. The honest answers are sentences like: “Formation lane in math, hint tutor on Tuesdays, Friday quiz closed.” “No student-facing chat; we talk about deepfakes in civics.” “One logged critique per month in history.” Those sentences travel. A shopping list of apps does not.

Next steps, in plain language

At the end of a term, write one page with four headings.

Solid. Skills with two unaided samples, dated, that a stranger could sample.

Next. One skill per subject, small enough for Chapter 5’s first problem.

Not yet for lane two. Units where error-detection is missing.

Hosting. Tools in use, host, off-switch test date, anything retired.

Send that page, plus three formation samples and at most one performance sample per subject, if someone asks what you did. A cloud of tabs is not the packet.

If you are preparing a high-school record for the outside world, name courses the way a registrar names them, then attach the skill list. “English 10” plus “unaided literary paragraph; sourced history of a claim; one logged critique of a model essay” is readable. “Completed AI literacy badge” is not.

A weekly fifteen-minute habit

Friday, after the exit ticket:

1. Title the week’s two or three artefacts and drop them in the folder.
2. Tick F or P on the one-page gradebook.
3. Write one next-move sentence.
4. Delete or archive any chat that contains a name, a school, or another child’s work.
5. Test the off-switch if a new tool appeared.

That is the whole records system. If Friday is impossible, monthly still beats a June panic.

A note on dual enrollment, umbrellas, and part-time campus courses. Their syllabus is their lane label. Keep a copy of it next to your F/P sheet so you do not fight their rules in the wrong folder. If they forbid tools, your formation samples *are* the course. If they require disclosure, your JCQ-style line is the course. If they grade only take-homes, still run unaided checks at home and keep them. You may need them when the grade and the skill diverge, which is the whole point of Chapter 3.

State-test weeks, if you have them, are formation weeks by definition. Record the prep as unaided. A helper the night before a test of unaided skill is rescue with a calendar.

Illustration, not a reported family: a parent keeps every chatbot printout and no unaided tickets. In June, the pile looks industrious. A co-op teacher asks what the student can do alone. The pile cannot answer. A thin folder with twelve titled formation samples can.

What “enough record” looks like

- A stranger can read three titles and know the lane, the skill, and the date.
- Each subject has unaided samples from the last month.

- Performance samples, if any, include log and oral note.
- No detector printouts.
- No required personal-agent homework in the trail.
- Placement sentences are about skills.
- Hygiene notes exist: host, training off, off-switch.
- The student can, once a term, pick a formation sample and redo a short version on demand.

You are done enough with records when the year is legible. You are not done when the drive is full.

If this still feels like bureaucracy, shrink it once more. Three formation samples per subject per term. Titles a stranger can read. One placement sentence. An off-switch test when something new appears. Everything else is optional furniture. The student who can redo a titled sample on demand has a record even if the printer dies. That redo is the point of Chapters 5 through 8, filed.

Store the folder where a fire or a broken laptop cannot erase the year: paper in a binder, or a dated copy on a drive you control. A transcript that lives only inside a product account is a transcript you do not host. Chapter 7 already told you who should host. The record is part of what is hosted, and it has to outlast a product login this year. Print the index once a term.

Resources next: named tools by fit — parent load, style, placement — so you can choose furniture that matches the house, without a sales pitch.

Chapter 10

Resources, named by fit

This list is furniture, not a ranking and not a store. Fit means three questions: parent load (how much of you the thing needs), style (chat, tutor, adaptive engine, agent, or adult helper), and placement (age, skill, formation or performance). A name appears because families meet it, or because a public system used it as a hosting model. Presence here is not a recommendation. Absence is not a slight.

Read Common Sense Media / Youth AI Safety Institute ratings before a child-facing login. Treat “unacceptable” as not-for-students. Treat “high risk” as not-for-students unless a school wrapper has actually changed the object. Prefer purpose-built education tools over general chat. Recheck when the model under the same brand name changes.

How to use the list this week

1. Name the job: retrieval practice, hint tutoring, adult planning, or judgment of a model paragraph.
2. Name the host (Chapter 7). If nobody can host it, stop.
3. Name the lane (Chapter 8). Formation tools wait until after the attempt. Performance tools are logged.
4. Match parent load to the week you actually have. A tool that needs you in the room is honest. A tool that promises a tutor while you are in another town is access, not tutoring.

Human high-dosage tutoring is still the benchmark if you are choosing between a person and a window. Nickow, Oreopoulos, and Quan pooled experimental preK–12 tutoring near three-tenths of a standard deviation. Earlier computer-assisted

practice typically sits lower. There is no What Works Clearinghouse report on generative tutors as of 30 August 2026. There is no identified learning trial on Hermes-class agents.

Adaptive systems (not chat, not agents)

These sequence items from a learner model. They do not browse the open web. They can run retrieval. They cannot replace the unaided ticket.

ALEKS, MATHia, i-Ready, DreamBox. Fit: mathematics practice, often already in a district or a homeschool purchase. Parent load: setup and a weekly look at the report; the session still needs an adult if engagement is thin. Style: adaptive engine. Placement: useful in a formation week as *practice*, not as the grade. Dedicated time can still produce little use; human presence raises use. Access is not tutoring.

Older intelligent tutoring systems (for example, Cognitive Tutor Algebra I) have What Works Clearinghouse reports. Those reports are not generative-AI results. Steal the habit of looking for independent reviews. A new chatbot does not inherit those reports.

Constrained tutors (hint-not-answer, logs)

Khanmigo (Khan Academy). Fit: Socratic math and related practice inside Khan Academy, with supervising adults able to see turns. Common Sense rated it low risk (August 2024), not zero. Parent load: you in the room; otherwise many students barely message. Style: tutor wrapper. Placement: middle-school practice block, attempt first. A two-year cluster trial in Hamilton County found small intention-to-treat effects that resembled practice without the chatbot; the median student messaged on about a third of practice days and in 17 percent of mistake sessions. A default-open window did not, in that message data, raise messaging. Use it as a hint ladder beside you, or skip it.

SchoolAI Spaces, MagicStudent, Raina, Dot. Fit: classroom-shaped, constrained assistants, not Grok Bot-class teammates. Parent load: teacher or parent configures the space; students should not meet an unconstrained persona. Style: tutor / classroom wrapper. Placement: assigned tasks in the room. Treat company-commissioned “critical thinking” numbers scored on chat logs as descriptions of the chat, not as unaided skill.

Rori. Fit: a WhatsApp math tutor studied in Ghana as extra time, about an hour a week, with the authors' own caution on a year-1 result. Parent load: a scheduled extra hour, not a teacher of record. Placement: extra practice in a system that already has classes. Extra time is extra time.

Custom and lab-shaped tutors (design is the treatment)

You will not buy these off a shelf as they ran in the studies. You can steal the shape.

Bastani's GPT Tutor. Same interface as unguarded GPT-4; full solutions withheld; teacher-written answers and common-mistake hints. Practice rose; the unaided exam matched control, while the unguarded arm later scored worse. Fit: the struggle policy of Chapter 5 in a prompt. Parent load: you write or collect the hints. Placement: high-school math practice, never as the exam.

Kestin's PS2 Pal. Instructor-written step-by-step physics in the prompt, sequential scaffolding, versus a strong active classroom, immediate post-test only. Fit: first-pass conceptual material. Not off-the-shelf ChatGPT. Delayed retention unmeasured.

Nigeria Copilot lab (De Simone and colleagues). After-school, science-of-learning prompts, teachers trained three days, students often paired. Fit: English practice in a supervised lab, hallucination taught as the student's job. Parent load: a person circulating. Placement: not Copilot alone.

LearnLM (DeepMind and Eedi, preprint). Expert tutors supervised every draft. Human-in-the-loop was literal. Fit: a reminder that supervision can be a gate on each message. Not a home product.

Tutor CoPilot (Wang and colleagues). Suggestions to the *tutor*, not a replacement. Fit: if a human tutor is already in the week, a whisper to them. About twenty dollars per tutor per year in that study. The adult stays in the conversation.

Mollick and Mollick's seven approaches (tutor, coach, mentor, teammate, student, simulator, tool) are a prompt menu. The Nigeria lab actually used science-of-learning prompts. You still choose the role. Default chat will choose "tool" and answer-giving "tutor."

Adult-facing helpers (the 2026 beachhead)

ChatGPT for Teachers. K–12 educator workspace, not intended for students. Fit: parent or teacher planning. Parent load: you own every output. Placement: adult evening, de-identified work, training off.

Claude for Teachers. Educator-only, eighteen-plus, scheduled Cowork, training off for verified teacher accounts, FERPA-aligned agreement on the education side. Fit: adult folder-scale jobs (for example, de-identified exit tickets). Parent load: review time; “in the loop” can add work. Placement: never as a child account. Prospect Schools gave feedback; a Detroit impact pilot was announced, not published as results.

OpenAI Workspace Agents. Research preview for eligible Edu workspaces, off by default, campus operations, humans review before send. Fit: registrar-scale adult workflows, not homework. No named campus production case was fetched in the research behind this book.

MagicSchool, Diffit. Teacher material generation — lesson ideas, leveled passages, questions. Fit: adult prep. Parent load: you still choose the mathematics and the reading level. MagicSchool’s ESSA Level III note is a positive association, company-funded, not a large independent trial of an agent. Diffit’s education-side claim is that it does not collect student data; read the current agreement.

Gemini in Workspace for Education / Gemini Notebook. Chat plus notebook on sources you provide, claimed Core-service data handling in the education editions. Fit: grounded notes on *your* sources. Style: not an autonomous school agent. Common Sense rated Gemini K-12 high risk (January 2026) on the student-facing side; privacy better, age-appropriate mental-health handling not. Keep child logins off unless a school host changes that object.

Parent load for all of the above: high on judgment, lower on typing. They return time only if you spend the saved minutes on struggle, orals, and noticing.

National wrappers (hosting models, not cart items)

Singapore Student Learning Space. Age bands: print-first in early primary, ministry-designed features under supervision later, then increasing ownership plus a duty to evaluate, check, and acknowledge. Learning Assistant redirects students who want to be spoon-fed. Fit: a public-law host with pedagogy designed in. Par-

ent load: none as a U.S. purchase; steal the sequence.

Estonia ITI. Socratic Estonian-language app that does not give answers; teachers get a paid ChatGPT or Gemini licence. Built to interrupt free-bot homework finishing. First-year review (August 2026): usage data not yet a basis to claim improved learning outcomes. Fit: the interrupt. Steal the split: adults get the general tool; students get the wrapper.

TeachAI toolkit; NYC traffic lights; Broward K–5 zero student-facing tools; Prince George’s AI addendum. Fit: questions to ask before anything touches a child. Parent load: a one-page copy of purpose, off-switch, no training on student work. Placement: every household that is the school of record.

Literacy programs and charts (vocabulary, not trials)

Cornell AI Critical Literacy Program. Four modules for incoming students; asks whether use contributes to learning. Pilot gains were program-reported, not a published randomized trial. Fit: a question to steal at the start of a year.

ISTELive 2026 six roles. Learner, Researcher, Synthesizer, Problem Solver, Connector, Storyteller. Conference profile, not an outcomes study. Fit: names for Chapter 8’s deeds, only if they sit on domain knowledge.

AI4K12 Five Big Ideas, grade-band charts. About AI first, then constrained tools. Fit: scope-and-sequence talk, not efficacy.

UNESCO student and teacher competency frameworks (2024); 2023 guidance. Human-centred; independent generative conversation from 13; untested tools not appropriate in primary. Fit: an age floor and a reminder that a ministry logo is not an implementation.

U.S. Department of Education 2023 report and 2026 supplemental priority. Humans in the loop; replacement rejected; grant scoring for literacy and tutoring, not a child-safety statute and not an agent mandate.

Assessment language (labels, not evidence)

Perkins et al., AI Assessment Scale. Shared words for levels of help. Communication tool. Later revision warns against using it merely as a prohibition ladder.

AFT full / partial / none. Teacher-decided labels; teachable-moment first response to misuse.

JCQ (from 30 April 2025). Own work; tool, URL, date; keep prompt and output; AI not the sole marker.

Harvard OUE three model policies. Forbid; allow with citation; assignment-specific.

Stanford CS221, Spring 2026. Collaborator, not answer key; no agentic repo-builder on the project; transcripts.

UC San Diego intro CS gate. Manual competency first, then Copilot — a customer story, not a cloud-agent trial. Steal the gate.

What these are not (keep the names)

Hermes 4 is an open-weight model family. **Hermes Agent** is an autonomous runtime (memory, browser, schedules, messaging). Nous does not recommend the larger Hermes 4 sizes inside the agent. Hermes Agent is not a child-safety school product. Optional home tutor for a family that wants it: not required, not fed with records, not a grade.

Grok the chat assistant is what Common Sense rated unacceptable in January 2026. **Grok Bot** (August 2026, with Cursor) has no education SKU; illustrated jobs are sales, inbox, expenses, bugs, deploys. **Grok-as-tutor in El Salvador** is a December 2025 announcement for the assistant across more than 5,000 schools, with no outcomes found as of 30 August 2026. Three objects. One name in ordinary talk. Keep them apart.

Character.AI, social companions, Meta AI. Companion class; Common Sense unacceptable. Not homework helpers. AFT would ban social-companion chatbots at least until 16.

Consumer ChatGPT / Claude / Gemini accounts. Style: chat. Fit: adult practice, or a performance-lane object to critique, hosted only if you can actually host them. High-risk teen ratings exist for ChatGPT-5 and Gemini K-12. Guardrails degrade in long conversations. Teens should not use them for mental health. Not FERPA school-officials in the abstract.

Cursor Cloud Agents, Claude Cowork, Codex Cloud. Agents for adult or developer work, repo-scoped, laptop-closed runtimes. Fit: staff and, at most, later

computing after a competency gate. Not a seventh-grader's homework.

Copilot (code). Worked examples at keystroke speed. Fit: after tracing exists. Novices shepherd or stray; weaker monitoring can be harmed.

A fit table you can copy

Job this week	Lower parent load	Higher parent load	Usually the wrong object
Retrieval of known facts	Adaptive practice engine, paper quiz	Oral retrieval with you	Chat that supplies the fact
New skill (formation)	Paper, worked example, faded hint card	You beside a hint tutor	Unguarded chat, overnight agent
Judgment (performance)	Printed model paragraph, pencil marks	Hosted tool, log, oral	Detector score, companion chat
Adult planning	Diffit / MagicSchool / teacher SKU, you edit	Scheduled Cowork on de-identified work	Child account "helping" you plan
Early elementary	Books, talk, literacy <i>about</i> AI	Adult-facing generation you review	Any student-facing chat
High-school practice	GPT-tutor <i>shape</i> , unaided Friday	You in the room, attempt gate	Take-home that needs a paid personal agent

Choose once, then stop shopping

Pick the smallest object that matches the job. A paper hint card is a complete formation tool. A hosted hint tutor is optional. An agent is not a tutor with more personality.

Re-audit annually or when the brand ships a new model. Turnitin's detector line and ChatGPT's version numbers have already moved under familiar names. Your off-switch test from Chapter 9 is the review.

If you want one civic add for the student: UNESCO's connectivity warning still applies at world scale: a chatbot is a poor substitute for missing teachers and missing electricity. In a U.S. kitchen the cousin of that warning is a purchase used as a substitute for a missing formation lane.

You now have struggle, judgment, a host, two lanes, a folder a stranger can read, and a list of furniture. The student still has to do the job. That was Chapter 1. It is still the job.

A Note on Sources

Studies that carry a mark in the chapters are listed in Notes, in one series at the back. That is where the full citations live, so the teaching pages can stay a teaching voice. Later chapters often point back in one sentence rather than reprint a note.

I did not invent a coefficient, a courtroom scene, or a product feature to fill a hole. Companion harm is named as litigation, not as a harm book. Detectors fail, and they fail unfairly — one plain sentence, then teach. There is no What Works Clearinghouse report on generative tutors as of 30 August 2026, and this book does not invent one.

The bibliography that follows is the working list from the research this book stands on. Access date 30 August 2026.

Bibliography

Autonomous AI and Education Michael Gannotti

Chicago Manual of Style, 17th edition, notes-bibliography system

Works named in the notes, plus the research list compiled 30 August 2026. Access date 30 August 2026 unless a later fetch is noted. Status tags after an entry ([Cited], [Consulted], [Vendor], [UNVERIFIED], and the rest) mark how the item was used in the research, not a ranking.

1. Statutes, regulations, executive orders, court filings

California. Assembly Bill 1159, *Student personal information* (2025–2026).

Amended 21 August 2026. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=20252026

Accessed 30 August 2026. Senate third reading / floor process as of that date;

not enacted. Would ban training generative AI on covered student information and would create HESIPA for higher education, proposed operative 1 July 2027.

[Cited]

California. K–12 Pupil Online Personal Information Protection Act (SOPIPA).

Cal. Bus. & Prof. Code § 22584. Operative 1 January 2016. https://leginfo.legislature.ca.gov/faces/codes_

Accessed 30 August 2026. No targeted ads, no noneducational profiles, no sale.

Pair with A.B. 1584 / Cal. Educ. Code § 49073.1. [Cited]

Children’s Online Privacy Protection Act of 1998. 15 U.S.C. §§ 6501–6506. Children’s Online Privacy Protection Rule, 16 C.F.R. pt. 312.

[https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-](https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa)

protection-rule-coppa. Accessed 30 August 2026. Under-13 personal information.

School-consent exception limited to the authorized educational purpose. [Cited]

European Commission. *Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU) 2024/1689*. C(2025) 5052 final. 29 July 2025. https://ai-act-service-desk.ec.europa.eu/sites/default/files/2025-08/guidelines_on_prohibited_artificial_intelligence_practices_established_by_regulation_eu_20241689_1.PDF. Accessed 30 August 2026. GPAI with emotion features must not be used by deployers in schools unless the medical/safety exception applies. Secondary sources also describe a 4 February 2025 draft approval; exact final-document date not re-fetched as PDF on the mid2 pass. [Cited]

European Commission. “AI Act | Shaping Europe’s Digital Future.” Page update 3 August 2026. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>. Accessed 30 August 2026. Application timeline including the Digital Omnibus. [Cited]

European Commission, AI Act Service Desk. “Annex III.” <https://ai-act-service-desk.ec.europa.eu/en/ai-act/annex-3>. Accessed 30 August 2026. Verbatim Annex III point 3 (education and vocational training). [Cited]

European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). In force 1 August 2024. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689>. Consolidated HTML used in the age-stages stream: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A02024R1689-20260727>. Accessed 30 August 2026. Art. 5(1)(f) and recital 44: emotion recognition in educational institutions prohibited from 2 February 2025 except medical/safety. Art. 4 AI-literacy duty from the same date. GPAI model duties from 2 August 2025. Transparency (chatbot disclosure) timed 2 August 2026. Annex III point 3: high-risk systems used to determine access to education, to evaluate learning outcomes including steering, to determine the appropriate level of education, or to monitor student behaviour during tests. Original stand-alone Annex III clock was 2 August 2026. [Cited]

European Union. Regulation (EU) 2026/1744 (Digital Omnibus). Political agreement 7 May 2026; OJ 24 July 2026; in force 27 July 2026. Mid2 fact-check treats this as **enacted** and as pushing stand-alone Annex III high-risk rules (including education access, evaluation, and test-monitoring) to **2 December 2027**; Annex I product-embedded high-risk later, 2 August 2028. Intimate-image / CSAM generation prohibition timed December 2026 via the Omnibus. **Confirm the consolidated EUR-Lex text of 2024/1689 as amended by 2026/1744 before typeset-**

ting. Do not silently keep 2 August 2026 as the only date. No completed Annex III education enforcement record. See GAPS. [Cited]

Exec. Order No. 14110, “Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,” 30 October 2023, 88 Fed. Reg. 75191 (1 November 2023). <https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>. Accessed 30 August 2026. Directed the Secretary of Education to develop education AI resources within 365 days. Rescinded 20 January 2025 by Exec. Order No. 14148. [Cited]

Exec. Order No. 14148, 20 January 2025. Cited in the research file as the initial-repeals order that rescinded EO 14110. Confirm against Federal Register text before quoting operative language. <https://www.federalregister.gov/documents/2025/01/20>. Accessed 30 August 2026. [Cited]

Exec. Order No. 14277, “Advancing Artificial Intelligence Education for American Youth,” 23 April 2025. <https://www.whitehouse.gov/presidential-actions/2025/04/advancing-artificial-intelligence-education-for-american-youth/>. Alternate: <https://www.presidency.ucsb.edu/documents/executive-order-14277-advancing-artificial-intelligence-education-for-american-youth>. Accessed 30 August 2026. Task Force, Presidential AI Challenge, public-private K-12 resources, ESEA/HEA Title II teacher-training priority, NSF, USDA 4-H. Adoption-and-workforce order, not a child-safety regulation. [Cited]

Family Educational Rights and Privacy Act. 20 U.S.C. § 1232g; 34 C.F.R. pt. 99. <https://www.ed.gov/privacy/ferpa>. Accessed 30 August 2026. School-official exception: 34 C.F.R. § 99.31(a)(1). Redisclosure: 34 C.F.R. § 99.33. Rights transfer at 18 or upon enrollment in postsecondary education. “FERPA compliant” is a contract-plus-configuration-plus-practice posture, not a product label. [Cited]

Federal Bureau of Investigation / Internet Crime Complaint Center. “Child Sexual Abuse Material Created by Generative AI and Similar Online Tools Is Illegal.” PSA, 29 March 2024. <https://www.ic3.gov/Media/Y2024/PSA240329>. Accessed 30 August 2026. [Cited]

Federal Trade Commission. “Children’s Online Privacy Protection Rule” (final amendments). 90 Fed. Reg. 16918 (22 April 2025). <https://www.federalregister.gov/documents/2025/04/2025-05904/childrens-online-privacy-protection-rule>. Accessed 30 August 2026. Commission vote 16 January 2025; effective 23 June 2025; general compliance 22 April 2026 except specified safe-harbor subsections. First major COPPA

Rule overhaul since 2013. Biometrics added to personal information. Separate verifiable parental consent for non-integral third-party disclosure. FTC: advertising, consideration, and “to train or otherwise develop artificial intelligence technologies” are not integral. Written retention policy; no indefinite retention; written security program. Proposed school-authorization / ed-tech exception **not finalized**, citing pending FERPA rulemaking RIN 1875-AA15 (NPRM unpublished as of the April 2025 preamble). [Cited]

Federal Trade Commission. “FTC Finalizes Changes to Children’s Privacy Rule Limiting Companies’ Ability to Monetize Kids’ Data.” Press release, January 2025. <https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-finalizes-changes-childrens-privacy-rule-limiting-companies-ability-monetize-kids-data>. Accessed 30 August 2026. [Consulted]

Federal Trade Commission. “FTC Launches Inquiry into AI Chatbots Acting as Companions.” 11 September 2025. <https://www.ftc.gov/news-events/news/press-releases/2025/09/ftc-launches-inquiry-ai-chatbots-acting-companions>. Order PDF: https://search.ftc.gov/system/files/ftc_gov/pdf/AICompanionChatbot6%28b%29Order.pdf. Accessed 30 August 2026. Section 6(b) study orders — not enforcement — to Alphabet, Character Technologies, Instagram, Meta, OpenAI, Snap, and xAI. Vote 3-0. [Cited]

Federal Trade Commission. *Policy Statement of the Federal Trade Commission on Education Technology*. https://search.ftc.gov/system/files/ftc_gov/pdf/Policy%20Statement%20of%20the. Accessed 30 August 2026. School authorization limited to the contracted educational service; no commercial use of school-collected data. [Cited]

Garcia v. Character Technologies, No. 6:24-cv-01903 (M.D. Fla.). Caption and docket number re-verified on the front fact-check pass, 30 August 2026. Sewell Setzer III, age 14, died 28 February 2024; suit October 2024 (AI Incident Database 826 records 22 October 2024). Senior U.S. District Judge Anne C. Conway, 21 May 2025, refused First Amendment dismissal at the pleading stage. Mediated settlement disclosed, court filing 7 January 2026. Allegations, a procedural ruling, and a disclosed settlement — not adjudicated causation. Do not reconstruct chats. [Cited]

Illinois. Student Online Personal Protection Act. 105 ILCS 85. Original P.A. 100-315 (2017); rewritten P.A. 101-516, effective 1 July 2021. <https://ilga.gov/legislation/ilcs/documents/010500850K15.htm>; prohibitions <https://ilga.gov/legislation/ilcs/documents/010500850K10.htm>. Accessed 30

August 2026. K-12. Written DPA before covered information transfers; no targeted ads, sale/rent, or non-school profiling. [Cited]

Individuals with Disabilities Education Act. 20 U.S.C. §§ 1400 et seq. Named in U.S. Department of Education, *Artificial Intelligence and the Future of Teaching and Learning* (May 2023), and in New York City Public Schools March 2026 AI guidance (IEP team, not AI). Full U.S.C. text not separately fetched. [Cited]

New York. Education Law § 2-d; 8 NYCRR pt. 121. Part 121 effective 29 January 2020. <https://codes.findlaw.com/ny/education-law/edn-sect-2-d/>. Accessed 30 August 2026. PII of students, teachers, and principals; Parents’ Bill of Rights; DPA in contracts; breach notice. [Cited]

Protection of Pupil Rights Amendment. 20 U.S.C. § 1232h; 34 C.F.R. pt. 98. <https://www.law.cornell.edu/uscode/text/20/1232h>. Accessed 30 August 2026. [Consulted]

Raine v. OpenAI (San Francisco Super. Ct., filed 26 August 2025). Adam Raine, age 16. Alleged harm after schoolwork use of ChatGPT, including method advice and note-drafting on GPT-4o as reported. OpenAI public line: redirected to crisis resources “more than 100 times”; reviewing the filing. Allegations, not a verdict. Do not reconstruct chats. [Cited]

Rehabilitation Act of 1973, § 504; Americans with Disabilities Act, Title II, including the 2024 web rule (WCAG 2.1 AA). Named in the manuscript (ch. 4 n. 4) with CAST UDL Guidelines 3.0. Full statute texts not separately fetched. [Cited]

Texas. Education Code § 32.155 (operator security and unique identifiers). https://texas.public.law/statutes/tex._educ._code_section_32.155. Accessed 30 August 2026. [Cited]

Texas. Responsible Artificial Intelligence Governance Act (TRAIGA), H.B. 149. Effective 1 January 2026. H.B. 3512 (AI training) and S.B. 1964 (AI systems and data management), effective 1 September 2025. <https://dir.texas.gov/technology-legislation>. Accessed 30 August 2026. DIR AI code of ethics; heightened scrutiny / NIST RMF; public-facing disclosure; applies to districts as local government. [Cited]

United Kingdom. Online Safety Act 2023. Named in UK Department for Education product-safety standards as applying to live-web / user-sharing generative AI. Full statute text not separately fetched. [Cited]

U.S. Department of Education. “Final Priority and Definitions—Secretary’s Supplemental Priority and Definitions on Advancing Artificial Intelligence in Education.” 91 Fed. Reg. 18774 (13 April 2026), effective 13 May 2026. Docket ED-2025-OS-0118. <https://www.federalregister.gov/documents/2026/04/13/2026-07087/final-priority-and-definitions-secretarys-supplemental-priority-and-definitions-on-advancing>. Accessed 30 August 2026. Discretionary-grant scoring tool, not a classroom mandate. AI defined via 15 U.S.C. § 9401(3). Does not define “agentic AI.” [Cited]

U.S. Department of Education, Student Privacy Policy Office. “Eligible Student.” <https://studentprivacy.ed.gov/content/eligible-student>. Accessed 30 August 2026. Rights transfer at 18 or postsecondary enrollment. [Cited]

U.S. Department of Education, Student Privacy Policy Office. “Responsibilities of Third-Party Service Providers under FERPA.” https://studentprivacy.ed.gov/sites/default/files/resource_documents/2025/08/Responsibilities_of_Third-Party_Service_Providers_under_FERPA.pdf. Accessed 30 August 2026. School-official exception four prongs; 34 C.F.R. § 99.31(a)(1)(i); direct control; authorized purposes only. [Cited]

2. Empirical research (journals, working papers, preprints labelled as such)

AICET / National University of Singapore. “The Educator as an Embedded AI Designer: Faculty Agency in GenAI Systems.” 2025. https://aicet.comp.nus.edu.sg/wp-content/uploads/2025/10/The_Educator_as_an_Embedded_AI_Designer_Faculty_Faculty_Agency_in.pdf. Accessed 30 August 2026. Intentional design, orchestrated mediation, reflective alignment. [Cited]

Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakçı, and Rei Mariman. “Generative AI without Guardrails Can Harm Learning: Evidence from High School Mathematics.” *Proceedings of the National Academy of Sciences* 122, no. 26 (2025): e2422633122. <https://doi.org/10.1073/pnas.2422633122>. Author PDF: https://hamsabastani.github.io/education_llm.pdf. Accessed 30 August 2026. Field RCT, large Turkish high school, Fall 2023-24, nearly 1,000 students, grades 9-11, four 90-minute sessions. ITT: GPT Base practice +0.137 (control mean 0.284) about +48 percent; unaided exam -0.054 (control mean 0.321) about -17 percent versus never-AI. GPT Tutor practice +0.361 about +127

percent; unaided exam -0.004, indistinguishable from control. Probe correctness about 51 percent. Students in the unguarded arm did not perceive the loss. Author-line spelling is Özge Kabakcı with cedilla on the PNAS page (correction notice 10.1073/pnas.2518204122). Full author PDF fetched. [Cited]

“Beyond Detection: Redesigning Authentic Assessment in an AI-Mediated World.” *Education Sciences* 15, no. 11 (2025): 1537. <https://doi.org/10.3390/educsci15111537>. Authenticity redesigned, not policed. Authors not listed in the research file. [Cited]

Bjork, Elizabeth L., and Robert A. Bjork. “Making Things Hard on Yourself, but in a Good Way: Creating Desirable Difficulties to Enhance Learning.” In *Psychology and the Real World*, edited by Morton Ann Gernsbacher et al., 56-64. New York: Worth, 2011. [Cited]

Bjork, Robert A. “Memory and Metamemory Considerations in the Training of Human Beings.” In *Metacognition*, edited by Janet Metcalfe and Arthur P. Shimamura, 185-205. Cambridge, MA: MIT Press, 1994. Origin of desirable difficulties. [Cited]

Bjork, Robert A., and Elizabeth L. Bjork. “Desirable Difficulties in Theory and Practice.” *Journal of Applied Research in Memory and Cognition* 9, no. 4 (2020): 475-79. Difficulty is desirable only when achievable. [Cited]

Bloom, Benjamin S. “The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring.” *Educational Researcher* 13, no. 6 (1984): 4-16. One-to-one *human* tutoring, not a language model. [Cited]

Chen, Ouhaio, Juan C. Castro-Alonso, Fred Paas, and John Sweller. “Extending Cognitive Load Theory to Incorporate Working Memory Resource Depletion.” *Educational Psychology Review* 30 (2018): 483-501. Cited in the manuscript (ch. 5 n. 8). The research JSON’s Chen 2018 DDF/CLT bridge note is incomplete; this is the title used on the page. [Cited]

Contractor, Zara, and Germán Reyes. “Experimental Evidence on the Learning Impact of Generative AI.” IZA Discussion Paper No. 18792. July 2026. <https://docs.iza.org/dp18792.pdf>. Accessed 30 August 2026. **Working paper, not a journal article.** Proctored RCT; unaided tests immediate and one week. Abstract: +0.27 SD on knowledge tests, persisting one week; delayed unaided essay style about 0.30 SD / relevance about 0.26 SD larger for augmentation than automation users. Full PDF timed out / partial; coefficients from abstract and IZA

page. [Cited; working paper; full PDF UNVERIFIED]

De Simone, Martin, Federico Tiberti, Maria Barron Rodriguez, Federico Manolio, Wuraola Mosuro, and E. J. Dikoru. “From Chalkboards to Chatbots: Evaluating the Impact of Generative AI on Learning Outcomes in Nigeria.” World Bank Policy Research Working Paper 11125. 2025. <https://documents1.worldbank.org/curated/en/099548105192529324/pdf/IDU-c09f40d8-9ff8-42dc-b315-591157499be7.pdf>. Accessed 30 August 2026. **Working paper.** RCT, 9 schools, Benin City. Copilot GPT-4 after-school English, teacher-guided science-of-learning prompts. ITT: composite 0.31 SD (SE 0.068); English 0.238 SD; regular third-term English exam 0.206 SD. Authors’ own Evans-Yuan conversion uses “0.24 SD.” Full PDF fetched. [Cited; working paper]

“Designing AI-Aware Assessment in Higher Education: A Cross-National Analysis of Generative AI Guidance and Design Patterns in English-Medium Universities.” 2026. <https://doi.org/10.17275/per.26.60.13.4>. Accessed 30 August 2026. 21 university guidance documents; four patterns (process portfolios, oral verification, critique of AI outputs, secure exams plus AI coursework). Authors not listed on the research landing. [Cited; authors omitted — see GAPS]

Deslauriers, Louis, Logan S. McCarty, Kelly Miller, Kristina Callaghan, and Greg Kestin. “Measuring Actual Learning versus Feeling of Learning in Response to Being Actively Engaged in the Classroom.” *Proceedings of the National Academy of Sciences* 116, no. 39 (2019): 19251-57. Illusion of learning; cited by Bastani on the perception mismatch. [Cited]

Diamond, Adele. “Executive Functions.” *Annual Review of Psychology* 64 (2013). <https://doi.org/10.1146/annurev-psych-113011-143750>. Working memory, inhibition, shifting; protracted development. Cited via secondary developmental reviews. [Cited]

Dietz, G., et al. “Theory of AI Mind: How Adults and Children Reason about the ‘Mental States’ of Conversational AI.” CogSci / Stanford SLL, 2023. https://sll.stanford.edu/docs/2023_cogsci/2023_Dietz_et_al_CogSci.pdf. Accessed 30 August 2026. Ages 3 to 8 treat AI like human agents (false belief). Full author line not locked from the research JSON. [Cited]

Dogan, S., U. Yazicilar Nalbantoglu, I. Celik, and N. Dogan. “Artificial Intelligence Professional Development: A Systematic Review of TPACK, Designs, and Effects for Teacher Learning.” *Professional Development in Education* 51, no. 3

(2025): 519-46. <https://doi.org/10.1080/19415257.2025.2454457>. Turkish characters simplified here; see GAPS. [Cited]

Eames, Taryn, Emma Brunskill, Bogdan Yamkovenko, Kodi Weatherholtz, and Philip Oreopoulos. "Computer-Assisted Learning in the Real World: How Khan Academy Influences Student Math Learning." *Proceedings of the National Academy of Sciences* 123, no. 10 (2026). Cited by Oreopoulos and Low: per-hour returns to Khan Academy without AI about 0.0047-0.0086 SD. Article number / pages not in the research file. [Cited]

Ennis, Robert H. "The Nature of Critical Thinking: Outlines of General Critical Thinking Dispositions and Abilities." 2011/2015. <https://criticalthinking.net/wp-content/uploads/2024/04/The-Nature-of-Critical-Thinking.pdf>. Accessed 30 August 2026. "Reasonable reflective thinking focused on deciding what to believe or do." [Cited]

Escueta, Maya, Andre J. Nickow, Philip Oreopoulos, and Vincent Quan. "Upgrading Education with Technology: Insights from Experimental Research." *Journal of Economic Literature* 58, no. 4 (2020): 897-996. Earlier computer-assisted learning typically 0.05-0.20 SD. [Cited]

"Evaluating Answer Leakage Robustness of LLM Tutors against Adversarial Student Attacks." ACL 2026 / arXiv:2604.18660. 2026. <https://arxiv.org/abs/2604.18660>. Accessed 30 August 2026. Manuscript cites the Zeng et al. line. Pedagogical tutors leak full answers under multi-turn student attacks. Fine-tuned adversarial student agent: example 7B tutor leakage 14 percent to 82 percent. Full author line not locked. [Cited; preprint or conference — see GAPS]

Facione, Peter A. *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction* (The Delphi Report). American Philosophical Association, 1990. <https://insightassessment.com/iaresource/the-delphi-report-a-statement-of-expert-consensus-on-the-definition-of-critical-thinking/>. Accessed 30 August 2026. Six core skills: interpretation, analysis, evaluation, inference, explanation, self-regulation. [Cited]

Facione, Peter A. *Critical Thinking: What It Is and Why It Counts*. Insight Assessment, 2015. <https://rtdw.ndm.edu/sites/default/files/pdf/Facione-2015.pdf>. Accessed 30 August 2026. [Cited]

Fan, Yizhou, L. Tang, H. Le, K. Shen, S. Tan, Y. Zhao, Y. Shen, X. Li, and D. Gašević. "Beware of Metacognitive Laziness: Effects of Genera-

tive Artificial Intelligence on Learning Motivation, Processes, and Performance.” *British Journal of Educational Technology* 56, no. 2 (2025): 489-530. <https://doi.org/10.1111/bjet.13544>. ERIC: <https://eric.ed.gov/?id=EJ1460793>. Accessed 30 August 2026. RCT N = 117 writing: ChatGPT vs human expert vs analytics vs none. Better essay scores; ns knowledge gain/transfer; fewer metacognitive processes. Publisher page Cloudflare-blocked; ERIC abstract used. Do not invent additional coefficients. [Cited; full BJET PDF UNVERIFIED]

“Generative AI, Cognitive Offloading, and Learner Agency in Higher Education: A Scoping Review.” *Behavioral Sciences* 16, no. 7 (2026): 1150. <https://doi.org/10.3390/bs16071150>. Accessed 30 August 2026. 123 studies. Authors unlisted on the DOI landing in the research file. Full text not fetched. [Consulted; full text UNVERIFIED]

Gerlich, Michael. “AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking.” *Societies* 15, no. 1 (2025): 6. Survey/correlational; not an RCT. Younger users more dependent. [Cited]

“Governing Generative AI in Higher Education: Emerging Policy Approaches and Support Ecosystems at Innovative U.S. Universities.” *International Journal for Educational Integrity* (2026). <https://doi.org/10.1007/s40979-026-00233-x>. AAC&U/Elon late-2024: 69 percent of leaders report written GenAI teaching policies. Yale/Cornell/MIT Sloan/USC caution against detectors as dispositive. Authors not listed in the research JSON. [Cited]

Hanham, Jose, Wayne Leahy, and John Sweller. “Cognitive Load Theory, Element Interactivity, and the Testing and Reverse Testing Effects.” *Applied Cognitive Psychology* 31, no. 3 (2017): 265-80. [Cited]

Harvard Undergraduate Survey on Generative AI. arXiv:2406.00833. 2024. <https://arxiv.org/abs/2406.00833>. <https://doi.org/10.48550/arXiv.2406.00833>. Accessed 30 August 2026. **Preprint survey**. Among AI users, 40 percent of students with no financial aid paid for AI products versus 20 percent with partial/full aid. Authors flag inequity if premium quality is real. Author line not in the research JSON. [Cited; preprint]

Henkel, Owen, Hannah Horne-Robinson, Neeve Kozhakhmetova, and Adam Lee. “Effective and Scalable Math Support: Experimental Evidence on the Impact of an AI Math Tutor in Ghana.” In *AIED 2024*, CCIS vol. 2150. Cham: Springer, 2024. <https://arxiv.org/pdf/2402.09809>. Accessed 30 August 2026. About 500 students; 1 hr/week WhatsApp Rori; Cohen’s $d = 0.36$; about \$5/student. Year-1

preliminary; authors flag caution. [Cited]

Hiebert, James, and Douglas A. Grouws. "The Effects of Classroom Mathematics Teaching on Students' Learning." In *Second Handbook of Research on Mathematics Teaching and Learning*, edited by Frank K. Lester Jr., 371-404. Charlotte, NC: Information Age, 2007. Productive struggle at p. 387. [Cited]

Institute of Education Sciences, What Works Clearinghouse. *Cognitive Tutor* intervention report. 2016. <https://ies.ed.gov/ncee/wwc/Intervention/818>. Accessed 30 August 2026. Older ITS reports exist. **No WWC intervention report specific to generative AI tutors** as of 30 August 2026. [Cited]

Kalyuga, Slava. "Expertise Reversal Effect and Its Instructional Implications." *Instructional Science* 35 (2007): 95-97. [Cited]

Kalyuga, Slava, Paul Ayres, Paul Chandler, and John Sweller. "The Expertise Reversal Effect." *Educational Psychologist* 38, no. 1 (2003): 23-31. [Cited]

Kapur, Manu. "Productive Failure in Learning Math." *Cognitive Science* 38, no. 5 (2014): 1008-22. [Cited]

Karpicke, Jeffrey D., and Henry L. Roediger III. "The Critical Importance of Retrieval for Learning." *Science* 319, no. 5865 (2008): 966-68. [Cited]

Kestin, Greg, Kelly Miller, Anna Klales, Timothy Milbourne, and Gregorio Ponti. "AI Tutoring Outperforms In-Class Active Learning: An RCT Introducing a Novel Research-Based Design in an Authentic Educational Setting." *Scientific Reports* 15 (2025): 17458. <https://doi.org/10.1038/s41598-025-97652-6>. Accessed 30 August 2026. Crossover RCT, Harvard Physical Sciences 2, Fall 2023, N = 194 eligible of 233 enrolled. Custom GPT-4 tutor ("PS2 Pal") versus the course's existing in-class active learning. Linear 0.63 SD; quantile 0.73-1.3 SD; median time 49 vs 60 minutes. Delayed retention unmeasured. Not off-the-shelf ChatGPT. Not an agent. Full article fetched. [Cited]

Khlaif, Zuheir N., et al. "Redesigning Assessments for AI-Enhanced Learning: A Framework for Educators in the Generative AI Era." *Education Sciences* 15, no. 2 (2025): 174. <https://doi.org/10.3390/educsci15020174>. Accessed 30 August 2026. Qualitative, 61 faculty; Against/Avoid/Adopt/Explore. Remaining author names not locked. [Cited]

Kosmyna, Nataliya, Eugene Hauptmann, Yeo T. Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, and Pattie Maes.

“Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task.” arXiv:2506.08872. 10 June 2025. <https://arxiv.org/abs/2506.08872>. MIT Media Lab: <https://www.media.mit.edu/publications/your-brain-on-chatgpt/>. Accessed 30 August 2026. **Preprint, not a journal article** as of 30 August 2026. N = 54 Boston-area adults (sessions 1-3); session 4 n = 18. EEG essay task; not K-12 and not agentic systems. Do not cite as population cognitive decline. Full PDF fetched. [Cited; preprint]

Kulik, James A., and J. D. Fletcher. “Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review.” *Review of Educational Research* 86, no. 1 (2016): 42-78. Pre-generative ITS. [Cited]

LearnLM Team, Google DeepMind, and Eedi. “AI Tutoring Can Safely and Effectively Support Students: An Exploratory RCT in UK Classrooms.” arXiv:2512.23633. 2025. <https://arxiv.org/html/2512.23633>. Accessed 30 August 2026. **Preprint**. N = 165 UK secondary; human-supervised LearnLM vs human tutor vs static hints. Session-level randomization; cumulative learning not identified; every message gated. Supervising-tutors-learned-Socratic-moves finding: interview evidence from n = 5, not a PD trial. Full HTML fetched. [Cited; preprint]

Li, K., P. Wang, and G. Chen. “How Can AI Be Integrated into Teacher Professional Development Programs? A Systematic Review Based on an Adapted Technology-Based Learning Model.” *Teaching and Teacher Education* 168 (2025): 105219. <https://doi.org/10.1016/j.tate.2025.105219>. 16 studies through 2024. Teacher knowledge/attitudes/self-efficacy; rarely student learning. [Cited]

Liang, Weixin, Mert Yuksekogonul, Yining Mao, Eric Wu, and James Zou. “GPT Detectors Are Biased against Non-Native English Writers.” *Patterns* 4, no. 7 (2023): 100779. <https://doi.org/10.1016/j.patter.2023.100779>. arXiv:2304.02819. Stanford HAI: <https://hai.stanford.edu/news/ai-detectors-biased-against-non-native-english-writers>. Accessed 30 August 2026. **61.22 percent** is the authors’ average false-positive rate on 91 human TOEFL essays (arXiv HTML); the *Patterns* HTML rounds the same average to 61.3 percent. This book keeps 61.22. Near-perfect on U.S. eighth-grade essays. **18/91** flagged by all seven detectors; **89/91 (97.8 percent)** flagged by at least one — the mid2 pass figure used on the page (introduction “ninety-seven percent” tightened to 97.8 in ch. 8). Perplexity metric penalizes constrained diction. Simple prompting bypasses detectors. [Cited]

Mollick, Ethan, and Lilach Mollick. “Assigning AI: Seven Approaches for Students, with Prompts.” arXiv:2306.10052. 2023. Tutor/coach/mentor/teammate/student/simulator/tool. Used as prompt source in De Simone et al. 2025. [Cited; preprint]

National Council of Teachers of Mathematics. *Principles to Actions: Ensuring Mathematical Success for All*. Reston, VA: NCTM, 2014. Warns against rescuing students too soon (p. 48 cited in secondary sources). [Cited]

Nickow, Andre, Philip Oreopoulos, and Vincent Quan. “The Promise of Tutoring for PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence.” *American Educational Research Journal* 61, no. 1 (2024): 74-107. <https://doi.org/10.3102/00028312231208687>. Pooled experimental preK-12 tutoring 0.288 SD (SE 0.029). [Cited]

“Not All Cognitive Offloading Is Equal: Distinguishing Dependent and Autonomous Offloading to Generative AI.” *Frontiers in Psychology* (2026). <https://doi.org/10.3389/fpsyg.2026.1878629>. Accessed 30 August 2026. Abstract-level: frequency of use is not the right variable; whether the user retains the evaluative loop is. Authors not listed in the research file. [Cited; abstract only]

Oreopoulos, Philip, Chloe R. Gibbs, Michael Jensen, and Joseph Price. “Teaching Teachers to Use Computer Assisted Learning Effectively: Experimental and Quasi-Experimental Evidence.” NBER Working Paper 32388. 2024. Forthcoming, *Journal of Human Resources*. Causal PD for CAL structure raises usage and achievement. Best analogue for generative-AI teacher PD; not a generative-AI trial. [Cited; working paper]

Oreopoulos, Philip, and Nina Low. “One Click Away: AI Tutoring with Khanmigo in a Two-Year School Experiment.” EdWorkingPaper 26-1551 / NBER Working Paper 35620. Annenberg Institute at Brown University, 2026. <https://doi.org/10.26300/kner-hv33>. PDF: <https://edworkingpapers.com/sites/default/files/ai26-1551.pdf>. NBER: <https://www.nber.org/papers/w35620>. Registry AEARCTR-0013519. Accessed 30 August 2026. **Working paper**. Cluster RCT, 18 Hamilton County, Tennessee middle schools; 6,902 student-terms; 53 clusters. ITT +1.26 NPR/term (SE 0.60); about 0.040 population SD/term; stacked about 0.06 SD/year, about 0.08 in year 2; implied full-year participation year 2 about 0.14 SD. Median messaged on one-third of practice days and in 17 percent of mistake sessions; 14.5 percent of messages mathematical. TCAP registered as co-primary, **not yet incorporated** as of the August 2026 paper. Cost about \$15 per student

per year. Full PDF fetched. Nina Low, not Nathaniel/Nathan. [Cited; working paper]

Paas, Fred, and Jeroen J. G. van Merriënboer. “Cognitive-Load Theory: Methods to Manage Working Memory Load in the Learning of Complex Tasks.” *Current Directions in Psychological Science* 29, no. 4 (2020): 394-98. [Cited]

Paas, Fred, and Jeroen J. G. van Merriënboer. “Variability of Worked Examples and Transfer of Geometrical Problem-Solving Skills.” *Journal of Educational Psychology* 86, no. 1 (1994): 122-33. [Cited]

Perkins, Mike, Leon Furze, Jasper Roe, and Jason MacVaugh. “The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment.” *Journal of University Teaching and Learning Practice* 21, no. 6 (2024). <https://doi.org/10.53761/q3azde36>. Shared vocabulary, not an RCT of learning. [Cited]

Perkins, Mike, Leon Furze, Jasper Roe, and Jason MacVaugh. “Reimagining the Artificial Intelligence Assessment Scale: A Refined Framework for Educational Assessment.” *Journal of University Teaching and Learning Practice* (2025/26). <https://doi.org/10.53761/rrm4y757>. Adds AI-exploration level; recasts as assessment-redesign framework. Full-text fetch timed out. [Cited]

Pew Research Center. “About a Quarter of U.S. Teens Have Used ChatGPT for Schoolwork – Double the Share in 2023.” 15 January 2025. <https://www.pewresearch.org/short-reads/2025/01/15/about-a-quarter-of-us-teens-have-used-chatgpt-for-schoolwork-double-the-share-in-2023/>. Accessed 30 August 2026. 26 percent of U.S. teens ages 13-17 (13 percent in 2023). Chat use, not agent use. Awareness rose most in the lowest-income households. 18 percent say essay-writing use is acceptable; 42 percent say not. A later February 2026 Pew chatbot survey exists; this book’s authority is the January 2025 ChatGPT number. [Cited]

Prather, James, Brent N. Reeves, Juho Leinonen, Stephen MacNeil, et al. “‘It’s Weird That It Knows What I Want’: Usability and Interactions with Copilot for Novice Programmers.” *ACM Transactions on Computer-Human Interaction* (2023). https://courses.cs.duke.edu/compsci290.2/spring24/readings/Copilot_for_Novice_Programmers.pdf. Accessed 30 August 2026. Shepherding and straying/drifting; metacognitive difficulties; overtrust. Volume/issue/pages not in the research file. The 2024 laboratory follow-on (greenfield assignments; 20 of 21 completed the task) is described in the pedagogy memo; a separate full bibliographic record was not

fetches. Do not treat either as a multi-site RCT of conceptual CS learning. [Cited]

Pyke, Aryn A., et al. "Does Difficulty Moderate Learning? A Comparative Analysis of the Desirable Difficulties Framework and Cognitive Load Theory." *Quarterly Journal of Experimental Psychology* (2025). <https://doi.org/10.1177/17470218241308143>. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12432286/>. Expertise and element interactivity moderate when difficulty is desirable. Remaining authors not locked. [Cited]

Risko, Evan F., and Sam J. Gilbert. "Cognitive Offloading." *Trends in Cognitive Sciences* 20, no. 9 (2016): 676-88. Canonical definition. [Cited]

Robinson, Carly D., David Gormley, Ana Trindade Ribeiro, and Susanna Loeb. "Access Is Not Enough: Human Support Improves Engagement with AI Tutoring." EdWorkingPaper 26-1451. Annenberg Institute, 2026. Cited in Oreopoulos and Low 2026. David Gormley and Ana Trindade Ribeiro (not Devin / Ana Teresa — that spelling belongs to Tutor CoPilot). [Cited; working paper]

Roediger, Henry L., III, and Jeffrey D. Karpicke. "Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention." *Psychological Science* 17, no. 3 (2006): 249-55. [Cited]

Soderstrom, Nicholas C., and Robert A. Bjork. "Learning versus Performance: An Integrative Review." *Perspectives on Psychological Science* 10, no. 2 (2015): 176-99. Conditions that raise performance can lower learning. Analytic frame; not re-fetched in full on 30 August 2026. [Cited]

Sparrow, Betsy, Jenny Liu, and Daniel M. Wegner. "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips." *Science* 333, no. 6043 (2011): 776-78. [Cited]

Stadler, Matthias, Maria Bannert, and Michael Sailer. "Cognitive Ease at a Cost: LLMs Reduce Mental Effort but Compromise Depth in Student Scientific Inquiry." *Computers in Human Behavior* 160 (2024): 108386. <https://doi.org/10.1016/j.chb.2024.108386>. Open PDF: <https://epub.ub.uni-muenchen.de/125262/1/1-s2.0-S0747563224002541-main.pdf>. RCT N = 91; ChatGPT-3.5 vs Google. Lower cognitive load, lower-quality reasoning/argumentation. [Cited]

"A Survey on Agentic AI-Powered Adaptive Learning Platforms." IEEE AIEI 2026. <https://doi.org/10.1109/aiei69164.2026.11497428>. Accessed 30 August

2026. Search snippet notes an empirical gap on fully agentic platforms. Authors not confirmed on fetch. Full text not fetched. [Cited; full text UNVERIFIED]

Sweller, John. “Cognitive Load during Problem Solving: Effects on Learning.” *Cognitive Science* 12, no. 2 (1988): 257-85. [Cited]

Sweller, John. “Cognitive Load Theory.” In *Psychology of Learning and Motivation*, vol. 55, edited by Jose P. Mestre and Brian H. Ross, 37-76. San Diego: Academic Press, 2011. Cited by Kestin et al. 2025. [Cited]

Sweller, John, Paul Ayres, and Slava Kalyuga. *Cognitive Load Theory*. New York: Springer, 2011. [Cited]

Teknum, Ryan, Roger Jin, Jai Suphavadeeprasit, Dakota Mahan, Jeffrey Quesnelle, Joe Li, Chen Guang, Shannon Sands, and Karan Malhotra. “Hermes 4 Technical Report.” arXiv:2508.18255. 25 August 2025. <https://arxiv.org/abs/2508.18255>. Accessed 30 August 2026. **Preprint**. Open-weight hybrid-reasoning LLM family. Distinguishes the model family from Hermes Agent. HTML conversion failed; existence confirmed via Hugging Face model card. **Quantitative benchmarks UNVERIFIED until the PDF is read**. [Cited; preprint; benchmarks UNVERIFIED]

Tervo-Clemmens, B., et al. “A Canonical Trajectory of Executive Function Maturation from Adolescence to Adulthood.” *Nature Communications* 14 (2023). <https://pubmed.ncbi.nlm.nih.gov/37903830/>. Significant EF growth ages 10-15; adult-like 18-20. N = 10,766. Remaining authors not locked. [Cited]

Trouche, Luc. “Mathematics Education and AI: Rethinking Instrumental Orchestration as the Structured Regulation of AI’s Pedagogical Agency.” French Institute of Education, ENS de Lyon, 2025. Lecture. AISIO: teacher regulates activation, delegation, validation of AI. Conceptual, not an RCT. The 2004 instrumental-orchestration original was not re-fetched. [Cited]

Trust-calibration work with upper-division students under extended Copilot use. arXiv:2509.13253. Named in ch. 3 n. 37. Authors and title not locked. Not a classroom RCT of conceptual learning. [Cited; incomplete record — see GAPS]

van Gog, Tamara, and John Sweller. “Not New, but Nearly Forgotten: The Testing Effect Decreases or Even Disappears as the Complexity of Learning Materials Increases.” *Educational Psychology Review* 27 (2015): 247-64. [Cited]

Van Vlasselaer, Marijke, Filip Van Droogenbroeck, and Bram Spruyt. “Who

Wrote This? Evaluating the Reliability of AI Detection Tools in Higher Education.” *International Journal for Educational Integrity* 22, art. 16 (2026). <https://doi.org/10.1007/s40979-026-00226-w>. Accessed 30 August 2026. 160 labelled documents. Turnitin 100 percent false negatives on fully-AI papers in that set (0-20 percent scores) against newer pipelines; humanised mixed text also poorly detected. Failure mode toward invisibility. Full PDF not re-read on the mid2 pass. [Cited]

Wang, Rose E., Ana T. Ribeiro, Carly D. Robinson, Susanna Loeb, and Dorottya Demszky. “Tutor CoPilot: A Human-AI Approach for Scaling Real-Time Expertise.” arXiv:2410.03017 / EdWorkingPaper. 2024. <https://arxiv.org/html/2410.03017>. Accessed 30 August 2026. RCT about 900 tutors, about 1,800 K-12 students. +4 pp topic mastery ITT; +9 pp for lower-rated tutors; about \$20/tutor/year. Keep Ana T. Ribeiro (this paper); do not collapse with Robinson 26-1451’s Ana Trindade Ribeiro. [Cited; preprint / working paper]

Wang, Shuai, et al. “The Efficacy of Artificial Intelligence-Enabled Adaptive Learning Systems from 2010 to 2022 on Learner Outcomes: A Meta-Analysis.” *Journal of Educational Computing Research* (2024). <https://doi.org/10.1177/07356331241240459>. N = 45 studies, g = 0.70 vs non-adaptive. Window ends 2022. **Not a generative-AI result.** Remaining authors not locked. [Cited]

Wang and Zhang. “Pedagogical Partnerships with Generative AI in Higher Education: How Dual Cognitive Pathways Paradoxically Enable Transformative Learning.” *International Journal of Educational Technology in Higher Education* (2026). <https://doi.org/10.1186/s41239-026-00585-x>. N = 912, three regions; U-shaped offloading curve. Full text not fetched; used at abstract/synthesis level. First names and outlet naming (IJETHE vs IJAIED) not locked. [Consulted; full text UNVERIFIED]

Warshauer, Hiroko K. “Productive Struggle in Middle School Mathematics Classrooms.” *Journal of Mathematics Teacher Education* 18 (2014): 375-400. [Cited]

Weber-Wulff, Debora, Alla Anohina-Naumeca, Sonja Bjelobaba, Tomáš Foltýnek, Jean Guerrero-Dib, Olumide Popoola, Petr Šigut, and Lorna Waddington. “Testing of Detection Tools for AI-Generated Text.” *International Journal for Educational Integrity* 19 (2023). <https://doi.org/10.1007/s40979-023-00146-z>. Commercial detectors neither accurate nor reliable enough for academic-integrity decisions. [Cited]

Willingham, Daniel T. “Critical Thinking: Why Is It So Hard to Teach?” *American*

Educator 31, no. 2 (Summer 2007): 8-19. <https://www.aft.org/sites/default/files/media/documents/2023/3> Accessed 30 August 2026. Verified 30 August 2026 against the AFT-hosted PDF. Exact bicycle sentence: “People who have sought to teach critical thinking have assumed that it is a skill, like riding a bicycle, and that, like other skills, once you learn it, you can apply it in any situation.” Processes of thinking “are intertwined with the content of thought (that is, domain knowledge).” Maxims can be taught, “but without background knowledge and practice, they probably will not be able to implement the advice they memorize.” “Literary thinking” in the book’s opening triad is a pedagogy-memo compression, not a third Willingham quotation. [Cited]

“Young Children’s Anthropomorphism of an AI-Powered Conversational Agent: Brain Activation and the Role of Parent Co-Presence.” *International Journal of Human-Computer Interaction* / arXiv 2512.02179. 2026. <https://doi.org/10.1080/10447318.2026.2678532>. Ages 5-6 perceptive and epistemic attribution; parent co-presence regulates; N = 23. Authors not locked in the research JSON. [Cited]

3. Policy, union, ministry, UNESCO / OECD / U.S. Department of Education

AI Leap Foundation / Estonian Ministry of Education and Research. TI-Hüpe / AI Leap. Announced February 2025 by President Alar Karis. <https://tihupe.ee/en/>. First-year review, 7 August 2026: <https://tihupe.ee/en/read-the-ai-leap-first-year-report/>. Accessed 30 August 2026. 154 of 156 upper secondary schools in 2025/26; teachers get paid ChatGPT or Gemini; about 20,000 grade 10-11 students get ITI, a Socratic Estonian-language app that does not give answers. Problem statement: 64-90 percent of students already using free commercial AI to finish homework. Public-private 50/50 funding. The 7 August 2026 review page **itself** states that first-year usage data “does not yet provide a basis for claiming that ITI has improved learning outcomes.” University of Tartu study (Jaan Aru; more than 9,000 students) expected later in 2026. Outcome coefficients not pulled. [Cited]

AI4K12 (AAAI / CSTA, NSF DRL-1846073). “Grade Band Progression Charts — Five Big Ideas in AI.” <https://ai4k12.org/gradeband-progression-charts/>. Ac-

cessed 30 August 2026. Working drafts, public. K-2 / 3-5 / 6-8 / 9-12 progression. [Cited]

American Academy of Pediatrics. “Digital Ecosystems, Children, and Adolescents: Policy Statement.” *Pediatrics* 157, no. 2 (2026). <https://doi.org/10.1542/peds.2025-075320>. Publisher HTML bot-blocked; content corroborated via AAP Center of Excellence explainer. Child-centered versus engagement-based design; 5 Cs. A dedicated AAP “AI in schools” policy does not yet exist. [Cited; publisher HTML UNVERIFIED]

American Academy of Pediatrics. “Digital Ecosystems, Children, and Adolescents: Technical Report.” *Pediatrics* 157, no. 2 (2026). <https://publications.aap.org/pediatrics/article/157/Ecosystems-Children-and-Adolescents>. Partial fetch. [Cited; partial]

American Academy of Pediatrics, Center of Excellence on Social Media and Youth Mental Health (Jenny Radesky). “Understanding the New AAP Digital Media Guidelines for Screen Time and Social Media.” 20 January 2026. <https://www.aap.org/en/patient-care/media-and-children/center-of-excellence-on-social-media-and-youth-mental-health/understanding-the-new-AAP-digital-media-guidelines/>. Accessed 30 August 2026. [Cited]

American Federation of Teachers. *Commonsense Guardrails for Using Advanced Technology in Schools*. Version 2. March 2025. https://www.aft.org/sites/default/files/media/documents/2025-03/Commonsense_Guardrails_for_Using_Advanced_Technology_in_Schools.pdf. HTML: <https://www.aft.org/reports/commonsense-guardrails-using-advanced-technology-schools>. Accessed 30 August 2026. Official title is “Using Advanced Technology in Schools,” **not** “for Artificial Intelligence in Education.” Value 1: student privacy and safety come first; the sacrifice clause is the **second** sentence (“The well-being and data security of students must be paramount, even if it means sacrificing some potential educational opportunities”). Value 3: educators choose the technology. Collective-bargaining annex: AI a mandatory subject; no surveillance; right to disconnect; equity funding. Value 9 environmental. Notes absence of federal regulation under the current administration. [Cited]

American Federation of Teachers. “On the Formation of a National Academy for AI Instruction.” Resolution adopted 22 July 2025. <https://www.aft.org/resolution/formation-national-academy-ai-instruction>. Press: <https://www.aft.org/press-release/aft-launch-national-academy-ai-instruction-microsoft-openai-anthropic-and-united>. Accessed 30 August 2026. \$23 million; Microsoft, OpenAI, Anthropic; UFT founding partner; NYC hub; free PD for 1.8 million members; **target** 400,000 educators / about 7.2 million students over five years — a plan, not a trained

headcount. Launch press is 8 July 2025; resolution adopted 22 July 2025. Do not mash with CNBC's ChatGPT for Teachers first-cohort about 150,000. [Cited]

American Federation of Teachers. "Preparing for the Age of Artificial Intelligence: Protecting Democracy, Workers, Students and the Planet." Resolution adopted 18 July 2026. <https://www.aft.org/resolution/preparing-age-artificial-intelligence-protecting-democracy-workers-students-and-planet>. Accessed 30 August 2026. Opposes student-facing AI in elementary; companion-chatbot ban until age 16; PK-2 screen limits except disability/compelling need. [Cited]

Broward County Public Schools. "Broward Powered by AI." Board Resolution No. 26-132, adopted 10 March 2026. <https://browardschools.ai/>. Accessed 30 August 2026. K-5: no district-approved student-facing AI tools. Grades 6-12: Khanmigo and HMH Writable; MagicStudent off. Copilot is staff (20,000 Copilot Premium staff licenses). No final grades by AI; no model training. Detectors not relied upon to make decisions about a student. SPARK cadre announced 18 August 2026. [Cited]

CAST. *UDL Guidelines 3.0*. 30 July 2024. <https://udlguidelines.cast.org/>. Accessed 30 August 2026. Learner agency; identity/bias; assistive tech 4.2. [Cited]

Center for Democracy and Technology. *Ableism and Disability Discrimination in New Surveillance Technologies*. May 2022. <https://cdt.org/wp-content/uploads/2022/05/2022-05-23-CDT-Ableism-and-Disability-Discrimination-in-New-Surveillance-Technologies-report-final-redu.pdf>. Accessed 30 August 2026. Remote proctoring flags disability-related behavior. Cited by NEA 2024. [Cited]

Council Rock School District. "Artificial Intelligence Guidelines." 2026. <https://resources.finalsite.net/images/v1757092961/crsdorg/j0eciv13o7ftocclwxma/AIGUIDELINES26.pdf>. Located via search; not fully extracted. [Consulted; full PDF not extracted]

EDUCAUSE. 2026 *EDUCAUSE: The Impact of AI on Learning Assessment Report*. Library page, June 2026. <https://library.educause.edu/resources/2026/6/2026-educause-the-impact-of-ai-on-learning-assessment-report>. Accessed 30 August 2026. Survey of 438 faculty/staff. Claims used from the library landing page only. **Full PDF not fetched.** [Cited; full PDF UNVERIFIED]

EDUCAUSE. 2026 *EDUCAUSE Horizon Report | Teaching and Learning Edition*. <https://library.educause.edu/resources/2026/5/2026-educause-horizon-report-teaching-and-learning-edition>. Accessed 30 August 2026. Listed

takeaway: process-based assessment. **Full PDF not fetched.** [Cited; full PDF UNVERIFIED]

Education International. “AI and Cognitive Offloading: Supporting Teachers to Shape AI’s Impact on Learning.” (Jason Lodge and Leslie Loble.) <https://www.ei-ie.org/en/item/32652:ai-and-cognitive-offloading-supporting-teachers-to-shape-ais-impact-on-learning>. Accessed 30 August 2026. Practitioner synthesis, not a primary experiment. [Cited]

Harvard FAS Office of Undergraduate Education. “Generative AI Guidance.” <https://oue.fas.harvard.edu/faculty-resources/generative-ai-guidance/>. Accessed 30 August 2026. Required Canvas policy; three model policies (forbid / allow with citation / assignment-specific). [Cited]

International Task Force on Teachers for Education 2030. *Promoting and Protecting Teacher Agency in the Age of Artificial Intelligence*. September 2025. <https://teachertaskforce.org/sites/default/files/2025-09/1149%5F25%5FPromoting%20and%20Protecting> Accessed 30 August 2026. Teachers irreplaceable; AI as augmentative. Full text not fully extracted. [Cited; partial]

ISTE. *ISTE Standards for Students*. Version 4.02, 2024. <https://iste.org/standards/students>. Accessed 30 August 2026. [Consulted]

ISTE Generation AI. “Profile of an AI-Ready Graduate.” <https://iste.org/generationai/news-updates/profile-of-an-ai-ready-graduate>. Accessed 30 August 2026. News-page summary of Culatta 2025 lists Ideator in place of Problem Solver; Culatta’s own ASCD headings match the ISTE Live 2026 six. Do not cite Culatta 2025 as an Ideator list. [Cited]

Joint Council for Qualifications. *AI Use in Assessments: Your Role in Protecting the Integrity of Qualifications*. Effective 30 April 2025. <https://www.jcq.org.uk/knowledge-hub/ai-use-in-assessments-your-role-in-protecting-the-integrity-of-qualifications-2/>. Accessed 30 August 2026. Own-work rule; acknowledgement = tool name + URL + date; non-editable prompt/output copy; AI cannot be the sole marker; detectors optional as one check. [Cited]

Little Falls Community Schools (MN). “Responsible Use of Artificial Intelligence” (Policy 625). 2026. <https://resources.finalsite.net/images/v1781805190/lfallsk12mnus/iooxv7omzintnuls> Located via search; H-AI-H model and grade-band access from snippets, not full extraction. [Consulted; full PDF not extracted]

Los Angeles Unified School District. “Guidelines for the Authorized Use of Artificial Intelligence for District” (BUL-151113). April 2024 framework. https://media.edlio.net/3476a301/6d998082/01503609/1dd0fae97a594f01aadad0a950d4ea11?_=BUL-151113_0_Guidelines_for_the_Authorized_Use_of_Artificial_Intelligence_for_District.pdf. Accessed 30 August 2026. Pre-vetting privacy/security; not a student-agent program. [Cited]

NAEYC. *Developmentally Appropriate Practice (DAP) Position Statement*. 2020. <https://www.naeyc.org/resources/position-statements/dap/contents>. Accessed 30 August 2026. Birth through 8. [Cited]

NAEYC. “Technology and Media in Early Childhood Education.” *Young Children*, Winter 2023. <https://www.naeyc.org/resources/pubs/yc/winter2023/technology-and-media-early-childhood-education>. Accessed 30 August 2026. [Cited]

NAEYC and Fred Rogers Center. *Technology and Interactive Media as Tools in Early Childhood Programs Serving Children from Birth through Age 8*. 2012. https://www.naeyc.org/sites/default/files/globally-shared/downloads/PDFs/resources/position-statements/ps_technology.pdf. Accessed 30 August 2026. Still the posted statement; revision underway 2025-26. Do not cite an updated NAEYC AI position that is not posted. [Cited]

National Education Association. *Report of the NEA Task Force on Artificial Intelligence in Education*. RA 2024. https://www.nea.org/sites/default/files/2024-06/report_of_the_nea_task_force_on_artificial_intelligence_in_education_ra_2024.pdf. Five principles: <https://www.nea.org/resource-library/artificial-intelligence-education/v-five-principles-use-artificial-intelligence-education>. Hub: <https://www.nea.org/ai>. Accessed 30 August 2026. Official order, verified 30 August 2026: (1) students and educators at the center; (2) evidence-based AI; (3) ethical development/use and strong data protection; (4) equitable access; (5) ongoing education with and about AI — literacy and agency. An earlier memo pass had swapped (4) and (5). No high-stakes AI-alone decisions. Detector bias; Texas A&M University-Commerce episode; Nevada funding algorithm; NY facial-recognition ban September 2023. COPPA NPRM status in the report is as of April 2024 — superseded by the 2025 final rule. [Cited]

New York City Public Schools. “Guidance on Artificial Intelligence.” March 2026. <https://www.schools.nyc.gov/about-us/vision-and-mission/artificial-intelligence/guidance-on-artificial-intelligence>. Accessed 30 August 2026. Traffic-light framework; ERMA; red list on grading, IEP/504, placement,

discipline, surveillance, counseling. December 2024 AI add-on: disclose AI capabilities; prohibit student-data model training; transparency. Chancellor's Regulation A-820. Grade-band student-use Playbook still forthcoming as of the fetched page (targeted June 2026). [Cited]

OECD. *OECD Digital Education Outlook 2026*. OECD Publishing, January 2026. https://www.oecd.org/en/publications/oecd-digital-education-outlook-2026_062a7394-en.html. PDF listed: https://www.oecd.org/content/dam/oecd/en/publications/reports/2026/digital-education-outlook-2026_940e0dd8/062a7394-en.pdf. Accessed 30 August 2026. Listing-page claim: unguided GenAI raises task performance without learning gains. TALIS 2024: 37 percent of teachers use GenAI for work — **listing-page figure; full PDF Cloudflare-blocked 30 August 2026**. Do not mash with EDUCAUSE 37 percent of *institutions* licensing chatbots. [Cited; full PDF UNVERIFIED]

OECD. “Policies Supporting Responsible and Systematic GenAI Adoption in Higher Education.” *OECD Education Spotlights* no. 23, 17 July 2026. <https://doi.org/10.1787/c4e5621f-en>. [Consulted]

Prince George's County Public Schools. *Artificial Intelligence Declaration Addendum*. <https://epi.pgcps.org/contentassets/0d463a5c451f40fcba26c2796d7a59c8/appendix-p-ai-declaration-addendum.pdf>. Accessed 30 August 2026. Specimen AI addendum: model disclosure, local vs cloud, FERPA/COPPA certification, no PII retained in generative models, no secondary use without district and parental approval, bias/harm audits on request, non-US subprocessors, feature off-switch. [Cited]

Quality Assurance Agency for Higher Education. *Reconsidering Assessment for the Chat GPT Era: QAA Advice on Developing Sustainable Assessment Strategies*. July 2023. <https://www.qaa.ac.uk/docs/qaa/members/reconsidering-assessment-for-the-chat-gpt-era.pdf>. Accessed 30 August 2026. [Cited]

Singapore Ministry of Education. “AI Usage in Schools” (parliamentary reply). 6 May 2026. <https://www.moe.gov.sg/news/parliamentary-replies/20260506-ai-usage-in-schools>. Accessed 30 August 2026. SLS Learning Assistant redirects students who “want to be spoon-fed with direct answers.” Designed tutor refusal, not an agent evaluation. [Cited]

Singapore Ministry of Education. “Artificial Intelligence in Education.” Page updated 31 July 2026. <https://www.moe.gov.sg/education-in-sg/educational-technology-journey/edtech-masterplan/artificial-intelligence-in-education>.

Accessed 30 August 2026. P1-3 print-first, no student chatbots; P4-6 only SLS education-designed tools under teacher supervision; secondary/pre-U integrity plus evaluation of AI output. [Cited]

Stanford University. “CS221: Artificial Intelligence — Generative AI Policy.” Spring 2026. <https://stanford-cs221.github.io/spring2026/>. Accessed 30 August 2026. Students may use GenAI as a human collaborator; may not ask for answers or to “check my work”; must share ChatGPT link or screenshots; “using an agentic coding tool such as Codex or Claude Code to build a repository from a project proposal is not allowed.” [Cited]

Student Data Privacy Consortium. *National Data Privacy Agreement* (NDPA) version 2. 2024. De-facto interstate DPA template with state exhibits. Not a statute. [Cited]

TeachAI (Code.org, ETS, ISTE, Khan Academy, World Economic Forum; advised by AASA, AFT, NEA, NSBA, CCSSO and others). *Foundational Policy Ideas for AI in Education / AI Guidance for Schools Toolkit*. 2024. <https://www.teachai.org/policy>; <https://www.teachai.org/toolkit-guidance>. Accessed 30 August 2026. Most copied U.S. K-12 template. [Cited]

UCLA Digital and Technology Solutions. “AI Use and Recommendation Guide.” <https://www.dts.ucla.edu/initiatives/ai/ai-use-recommendation-guide>. Accessed 30 August 2026. BruinLearn AI hub; OpenAI agreement first CA university at scale (vendor/admin claim); no FERPA/HIPAA in public tools. [Cited]

UCLA Teaching and Learning Center. “Using Generative AI Reflectively and Responsibly in Teaching and Learning.” <https://teaching.ucla.edu/resources/teaching-guides/using-generative-ai-reflectively-and-responsibly-in-teaching-and-learning/>. Accessed 30 August 2026. UCUES 2024: 67 percent of UCLA undergraduate *respondents* used AI tools at least several times a year; top uses brainstorming, researching a topic, and studying for exams. SAIRO brief: 4,706 respondents, 15 percent response rate. <https://sairo.ucla.edu/file/69a9752b-ba55-46a1-8cf1-e2a963ffbb97>. [Cited]

UK Department for Education. *Generative AI: Product Safety Standards*. First published 22 January 2025; updated 19 January 2026. <https://www.gov.uk/government/publications/generative-ai-product-safety-standards/generative-ai-product-safety-standards>. Accessed 30 August 2026. Most operational English-language child-safety spec: filtering that holds for whole conversations; jailbreak resistance; no anthropomorphism; no sycophancy/dark patterns; hard time limits; mental-health pathways to hu-

mans; cognitive-offloading detection; no commercial training on inputs without permission. Live-web/search GenAI under the Online Safety Act 2023. [Cited]

UK Department for Education. *Generative Artificial Intelligence (AI) in Education*. First published 29 March 2023; updated 12 August 2025. <https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education/generative-artificial-intelligence-ai-in-education>. Accessed 30 August 2026. Applies to England. Teacher-facing lower risk; pupil-facing only with safeguards, KCSIE, filtering/monitoring, age limits. Do not train models on student original work without permission. [Cited]

UNESCO. “AI and Education: Protecting the Rights of Learners.” 2024. <https://www.unesco.org/en/articles/ai-and-education-protecting-rights-learners>. Accessed 30 August 2026. About 2.6 billion people without internet (2024 figure UNESCO still uses); about 1 in 4 primary schools without electricity; about 60 percent of primary schools unconnected. [Cited]

UNESCO. “Artificial Intelligence in Education.” <https://www.unesco.org/en/digital-education/artificial-intelligence>. Accessed 30 August 2026. Human-centred framing; “preserving human agency in a world of automation.” [Cited]

UNESCO. *Beijing Consensus on Artificial Intelligence and Education*. 2019. <https://unesdoc.unesco.org/ark:/48223/pf0000368303>. [Consulted]

UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. 2021. <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>. First global AI ethics instrument. [Consulted]

UNESCO. “UNESCO: Governments Must Quickly Regulate Generative AI in Schools.” 7 September 2023. <https://www.unesco.org/en/articles/unesco-governments-must-quickly-regulate-generative-ai-schools>. Accessed 30 August 2026. Under 10 percent of 450-plus surveyed schools/universities had GenAI policies. [Cited]

UNESCO. “What You Need to Know about UNESCO’s New AI Competency Frameworks for Students and Teachers.” 2024. <https://www.unesco.org/en/articles/what-you-need-know-about-unescos-new-ai-competency-frameworks-students-and-teachers>. Accessed 30 August 2026. [Cited]

UNESCO (Fengchun Miao and Kelly Shiohira). *AI Competency Framework for Students*. 8 August 2024. <https://www.unesco.org/en/articles/ai-competency-framework-students>. 12 competencies / 4 dimensions / Understand-Apply-Create.

[Cited]

UNESCO (Fengchun Miao and Mutlu Cukurova). *AI Competency Framework for Teachers*. 8 August 2024. <https://www.unesco.org/en/articles/ai-competency-framework-teachers>. 15 competencies / 5 dimensions / Acquire-Deepen-Create. [Cited]

UNESCO (Fengchun Miao and Wayne Holmes). *Guidance for Generative AI in Education and Research*. 7 September 2023. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>. Accessed 30 August 2026. Age 13 for independent GenAI conversation; under-13 / primary use of untested tools “not appropriate”; seven governmental steps; Global North training-data bias. Last UNESCO page update noted 16 January 2026. [Cited]

University of Illinois Chicago, Center for the Advancement of Teaching Excellence. “Adapting Assignments for the Age of AI.” <https://teaching.uic.edu/cate-teaching-guides/digital-learning/adapting-assignments-for-the-age-of-ai/>. Accessed 30 August 2026. [Cited]

U.S. Department of Education. *Navigating Artificial Intelligence in Postsecondary Education*. 2025. ERIC ED670768. <https://files.eric.ed.gov/fulltext/ED670768.pdf>. Accessed 30 August 2026. [Cited]

U.S. Department of Education. “U.S. Department of Education Issues Guidance on Artificial Intelligence Use in Schools, Proposes Additional Supplemental Priority.” Press release, 2025. <http://www.ed.gov/about/news/press-release/us-department-of-education-issues-guidance-artificial-intelligence-use-schools-proposes-additional-supplemental-priority>. Accessed 30 August 2026. [Consulted]

U.S. Department of Education / ERIC ED661924. *Empowering Education Leaders* [AI toolkit]. 2024. <https://files.eric.ed.gov/fulltext/ED661924.pdf>. Accessed 30 August 2026. Responsive to Biden EO 14110 365-day clock. District use policies; opt-out of AI-enabled applications. [Cited]

U.S. Department of Education, Office of Educational Technology. *Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations*. May 2023. <https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf>. Accessed 30 August 2026. Seven recommendations. Humans in the loop. The Department “firmly rejects the idea that AI could replace teachers.” FERPA/IDEA already apply. Public domain. [Cited]

U.S. Department of Education, Office of Educational Technology. *Designing for Education with Artificial Intelligence: An Essential Guide for Developers*. 2024. ERIC ED661949. <https://files.eric.ed.gov/fulltext/ED661949.pdf>. Accessed 30 August 2026. [Cited]

U.S. Department of Education, OPEPD. “Guidance on the Use of Federal Grant Funds to Improve Education Outcomes Using Artificial Intelligence (AI).” Dear Colleague letter, 22 July 2025. <https://www.ed.gov/media/document/oepd-ai-dear-colleague-letter-7222025-110427.pdf>. Accessed 30 August 2026. Allowable uses for tutoring, advising, materials; human-oversight condition. [Cited]

World Economic Forum. “AI Won’t Replace Teachers, Says This Global Union.” Interview with David Edwards, Education International. July 2024. <https://www.weforum.org/stories/2024/07/artificial-intelligence-education-teachers-union/>. Accessed 30 August 2026. Relational not transactional. Equity risk: human teachers as a luxury of the rich, bots for the poor. [Cited]

4. Product documentation (labelled vendor)

Anthropic. “Introducing Claude for Teachers.” 14 July 2026 (updated 21 July and 28 August 2026). <https://www.anthropic.com/news/claude-for-teachers>. Product: <https://claude.com/solutions/teachers>. Accessed 30 August 2026. **Vendor.** Educator-only; Claude 18+; FERPA-aligned DPA; training off for verified teacher accounts. Prospect Schools (Brooklyn) feedback; Detroit Public Schools Community District announced impact pilot, results unpublished. Sign-up by 30 June 2027 for a year of free access. Cowork scheduling; Learning Commons. Connectors named as ecosystem, not Anthropic-built. [Cited; vendor]

Character.AI. “An Update on Changes to Our Under-18 Experience.” October/November 2025. <https://blog.character.ai/an-update-on-changes-to-our-under-18-experience/>. Accessed 30 August 2026. **Vendor.** Open-ended character chat removed for under-18s from 24-25 November 2025. [Cited; vendor]

Cursor. “Cloud Agents.” <https://cursor.com/docs/cloud-agent>. Accessed 30 August 2026. **Vendor.** Isolated Cursor-managed Linux VMs; PR artifacts; kickoff surfaces. Formerly named Background Agents. [Cited; vendor]

Diffit. “Diffit Plans.” <https://web.diffit.me/plans>. App: <https://app.diffit.me/>. Ac-

cessed 30 August 2026. **Vendor.** FERPA/COPPA; vendor states it does not collect student data. [Cited; vendor]

Google. “Google and Khan Academy Build Classroom AI Tools.” <https://blog.google/products-and-platforms/products/education/khan-academy-back-to-school/>. Accessed 30 August 2026. **Vendor.** Gemini-powered Khanmigo features for the 2026 school year. [Cited; vendor]

Google Workspace Help. “Quickstart Guide to Gemini and Gemini Notebook for Education.” <https://support.google.com/a/answer/16350447>. Accessed 30 August 2026. **Vendor/admin.** Free in all Education editions; Core-service privacy claims. LearnLM marketing language labelled as such. Not a documented autonomous-agent SKU for schools. [Cited; vendor]

Khan Academy. “Becoming a Khan Academy Districts Partner.” <https://blog.khanacademy.org/becoming-a-khan-academy-districts-partner/>. Accessed 30 August 2026. **Vendor.** \$15/student Khanmigo; Clever or ClassLink; minimum 250 licenses; grades 3-12. “Over 400 partners across the US” is a vendor claim. Named partner: School City of Hobart, Indiana. [Cited; vendor]

Khan Academy. “Khan Academy Districts 2026-27 Renewal Guide.” <https://blog.khanacademy.org/khan-academy-districts-2026-27-renewal-guide/>. Accessed 30 August 2026. **Vendor.** Platform reimagining timeline. Not an efficacy study. [Consulted; vendor]

Magic School, Inc. “Midway ISD Case Study.” 2025. <https://www.magicschool.ai/case-studies/midway-isd>. Accessed 30 August 2026. **Vendor case study.** Midway ISD, Texas (about 8,700 students); 630-plus teachers trained summer 2024; 23-page DPA, no redlines. Not independent evaluation. [Cited; vendor]

Magic School, Inc. “Trust, Data Protection and Privacy for Schools.” <https://www.magicschool.ai/privacy/trust>. Accessed 30 August 2026. **Vendor.** iKeepSafe FERPA/COPPA; SOC 2 Type II; ESSA Level III (Instructure) — vendor-funded evidence tier, not a large independent RCT. [Cited; vendor]

Microsoft Customer Stories. “UC San Diego Prepares Students for AI-Driven Industry with GitHub Copilot.” <https://www.microsoft.com/en/customers/story/26539-university-of-california-san-diego-github-copilot>. Accessed 30 August 2026. **Vendor customer story.** Porter and Shah: manual competency then Copilot. Not a peer-reviewed trial. [Cited; vendor]

Nous Research. “Hermes 4 — Qwen-3 14B” (model card). 2025. <https://huggingface.co/NousResearch/Hermes4>

4-14B-FP8. Accessed 30 August 2026. **Vendor.** Distinguishes the LLM family from the agent runtime. Cites arXiv 2508.18255. [Cited; vendor]

Nous Research. Hermes Agent. <https://hermes-agent.org/>. Docs: <https://hermes-agent.nousresearch.com/docs/>. GitHub: <https://github.com/nousresearch/hermes-agent/>. Nous Portal: <https://hermes-agent.nousresearch.com/docs/integrations/nous-portal>. Research-paper-writing skill: <https://hermes-agent.nousresearch.com/docs/user-guide/skills/bundled/research/research-research-paper-writing>. Accessed 30 August 2026. **Vendor.** Open-source MIT-licensed autonomous runtime. Vendor site dates the launch to **February 2026 without a day**; 25 February is the first-release date recorded in this book’s responsible-use stream (GitHub/releases), not a date printed on the vendor homepage. “Not a chatbot. Not a copilot. An agent that lives on your machine.” No telemetry claimed; data in `~/hermes/`. **Not a child-safety certified school product.** No institutional AUP/DPA/SOP found. Hermes 4 (70B / 405B) is “not recommended for use inside Hermes Agent.” Recommended agent models on the Portal page include Claude Sonnet 4.6, GPT-5.5 Pro, google/gemini-3-pro-preview, DeepSeek V4 Pro. The **agent** installs locally; **open-weight Hermes models** can be served on llama.cpp, MLX, vLLM, Ollama, or LM Studio — do not conflate. [Cited; vendor]

OpenAI. “A Free Version of ChatGPT Built for Teachers.” 19 November 2025. <https://openai.com/index/chatgpt-for-teachers/>. Accessed 30 August 2026 (re-fetched on the mid2/front fact-check pass). **Vendor.** Free through June 2028 on this page; first cohort “nearly 150,000 teachers and staff.” A separate expansion URL timed out on the landscape pass. [Cited; vendor]

OpenAI. “Bringing ChatGPT for Teachers to More U.S. School Districts.” <https://openai.com/index/bringing-chatgpt-for-teachers-to-more-us-school-districts/>. FETCH TIMED OUT 30 August 2026. Additional 2026 district counts UNVERIFIED. [Cited as timed-out; extra counts UNVERIFIED]

OpenAI. “Introducing ChatGPT Edu” / business data privacy / enterprise privacy. <https://openai.com/index/introducing-chatgpt-edu/>; <https://openai.com/business-data/>; <https://openai.com/enterprise-privacy/>. Accessed 30 August 2026. **Vendor claim:** ChatGPT Edu/Enterprise/Business/ Teachers/API data not used to train models by default. Not a FERPA determination. [Cited; vendor]

OpenAI. “Introducing ChatGPT for Teens: Built for Learning, Backed by Protections.” 2026. <https://openai.com/index/chatgpt-for-teens/>. Accessed 30 August 2026. **Vendor.** Under-18 model spec; no romantic dependence language. [Cited;

vendor]

OpenAI. “Our Approach to Age Prediction.” 2026. <https://openai.com/index/our-approach-to-age-prediction/>. Accessed 30 August 2026. **Vendor**. Behavioral age signals; default to teen protections. [Cited; vendor]

OpenAI. “Terms of Use.” Effective 1 January 2026. <https://openai.com/policies/row-terms-of-use/>. Accessed 30 August 2026. **Vendor**. Minimum age 13; parent permission under 18. [Cited; vendor]

OpenAI Academy. “Deploying Codex in Higher Education.” 9 April 2026. <https://academy.openai.com/public/clubs/higher-education-05x4z/blogs/deploying-codex-in-higher-education-2026-04-09>. Accessed 30 August 2026. **Vendor playbook**. App vs Cloud vs CLI; security/repo scoping. Not an outcomes study. [Cited; vendor]

OpenAI Academy. “How to Build a Workspace Agent for Higher Education.” 6 May 2026. <https://academy.openai.com/en/public/clubs/higher-education-05x4z/blogs/how-to-build-a-workspace-agent-for-higher-education-2026-05-06>. [Consulted; vendor]

OpenAI Academy (Kirk Gulezian). “Understanding Workspace Agents in Higher Education.” 23 April 2026. <https://academy.openai.com/public/clubs/higher-education-05x4z/blogs/understanding-workspace-agents-higher-education>. Accessed 30 August 2026. **Vendor**. Research preview; off by default; admin-gated; humans review before anything is sent. Hypothetical campus workflows, not named-campus case studies. [Cited; vendor]

SchoolAI. “Give Every Student a Personalized Learning Experience | SchoolAI Spaces.” <https://schoolai.com/products/spaces>. Accessed 30 August 2026. **Vendor**. Teacher-designed Spaces. [Consulted; vendor]

SchoolAI. “SchoolAI Makes Students Think.” <https://schoolai.com/research/schoolai-makes-students-think>. Accessed 30 August 2026. **Vendor research**. Jordan School District (UT), 82 teachers, 13,882 conversations, 28 percent critical-thinking claim, ESSA Level III. Vendor-commissioned; not an RCT; Bloom rubric on chat logs. [Cited; vendor]

xAI. “AI Teammates That Finish the Work | Grok Bot.” <https://x.ai/bot>. Use cases: <https://x.ai/bot/use-cases>. Accessed 30 August 2026. **Vendor**. Distinguishes Bots from AI assistants (own computer, 24/7, parallel). No education SKU. [Cited; vendor]

xAI. “Grok Bot Is Now Included with More Plans.” 26 August 2026. <https://x.ai/news/grok-bot-more-plans>. Accessed 30 August 2026. **Vendor.** Expansion to SuperGrok plans and Cursor Pro/Teams. Website builder/deploy and meeting stand-in are on this expansion post, not the 11 August launch. None educational. [Cited; vendor]

xAI. Grok Bot FAQ. <https://docs.x.ai/grok-bot/faq>. Landscape fetch timed out; **re-fetched** on the front-matter fact-check pass, 30 August 2026. Requires cloud data storage; Legacy Privacy Mode not supported; training opt-out follows Cursor account settings. Still no schools, students, or teachers. Still no education SKU, DPA, or classroom SOP. End-errata did not re-clear this FAQ; front-errata is the later same-day fetch. [Cited; vendor]

xAI. “Introducing Grok Bot.” 11 August 2026. <https://x.ai/news/introducing-grok-bot>. Accessed 30 August 2026. **Vendor.** Early beta. “Bots have their own computer... keep working 24/7.” Sales outbound, CRM, inbox, expenses, bug reproduction on the launch post. No education SKU; product pages do not mention schools, students, or teachers. [Cited; vendor]

xAI. “Welcome to Grok” (docs overview). <https://docs.x.ai/grok/overview>. Accessed 30 August 2026. **Vendor.** Distinguishes Grok chat product from Grok Bot. [Cited; vendor]

xAI. “xAI and El Salvador Pioneer the World’s First Nationwide AI Education Program.” 11 December 2025. <https://x.ai/news/el-salvador-partnership>. Accessed 30 August 2026. **Vendor announcement.** Grok (the assistant, not Grok Bot) across more than 5,000 public schools over two years; “personalized learning to over one million students.” Announced national program, not a documented completed deployment; **no outcome evaluation found** as of 30 August 2026. [Cited; vendor]

5. Journalism and ratings

AI Incident Database. “Incident 826: Character.ai Chatbot Allegedly Influenced Teen User Toward Suicide.” <https://incidentdatabase.ai/cite/826/>. Accessed 30 August 2026. Timeline: use from April 2023; death 28 February 2024; suit 22 October 2024; Google/Character mediated settlement disclosed 7 January 2026.

[Cited]

American Enterprise Institute. “Can an AI-Powered Tutor Produce Meaningful Results?” Interview with Kristen DiCerbo (Khan Academy). <https://www.aei.org/commentary/can-an-ai-powered-tutor-produce-meaningful-results/>. Accessed 30 August 2026. Partner-district users about 68,000 (2023-24) to more than 700,000 (2024-25); districts 45 to 380-plus — **vendor claims**, not efficacy RCTs. [Cited]

Associated Press. “AI Chatbot Pushed Teen to Kill Himself, Lawsuit Alleges.” <https://apnews.com/article/chatbot-ai-lawsuit-suicide-teen-artificial-intelligence-9d48adc572100822fdb3c90d1456bd0>. Companion on the Conway ruling: <https://apnews.com/article/ai-lawsuit-suicide-artificial-intelligence-free-speech-ccc77a5ff5a84bda753d2b044c83d4b6>. Accessed 30 August 2026. Allegations. [Cited]

Beduya, Jose. “Cornell Expands AI Critical Literacy Program for Incoming Students.” *Cornell Chronicle*, August 2026. <https://news.cornell.edu/stories/2026/08/cornell-expands-ai-critical-literacy-program-incoming-students>. Accessed 30 August 2026. Four Canvas modules: What is Generative AI?; Ethical Questions Surrounding Generative AI; GenAI and Learning; Creating an AI Policy for Personal Use. CTI-reported literacy gains, not a published RCT. [Cited]

Burns, Mary. “What the Research Shows about Generative AI in Tutoring.” Brookings Institution, 27 January 2026. <https://www.brookings.edu/articles/what-the-research-shows-about-generative-ai-in-tutoring/>. Accessed 30 August 2026. Think-tank commentary. ITS vs genAI-tutor distinction. Not itself an RCT. [Cited]

CBS News / 60 Minutes. “How Classroom AI Khanmigo Can Help Students in Emotional Distress.” 8 December 2024. <https://www.cbsnews.com/news/how-classroom-ai-khanmigo-can-help-students-in-emotional-distress-60-minutes/>. Accessed 30 August 2026. Vendor claims: teacher-visible logs; self-harm flags; OpenAI contractually not training on Khanmigo data. Claims, not an independent audit. [Cited]

Chalk, Dan. “Midland Public Schools Selects SchoolAI for Grades 6-12.” *Our-Midland*, 19 August 2026. <https://www.ourmidland.com/news/article/midland-schools-school-ai-22393016.php>. Accessed 30 August 2026. Board selection, not outcomes. [Cited]

Chalkbeat New York. “What NYC’s New AI School Rules Say, and What Still Re-

mains Unclear.” 24 March 2026. <https://www.chalkbeat.org/newyork/2026/03/24/preliminary-ai-policy-nyc-schools/>. Accessed 30 August 2026. [Cited]

Clemson News. “GPT-5.5, Codex and Clemson-Hosted Models Give Researchers New Options for AI-Assisted Work.” <https://news.clemson.edu/gpt-5-5-codex-and-clemson-hosted-models-give-researchers-new-options-for-ai-assisted-work/>. Accessed 30 August 2026. Institutional access, not learning outcomes. [Cited]

CNBC. “Cursor Announces Major Update as AI Coding Agent Battle Heats Up.” 24 February 2026. <https://www.cnbc.com/2026/02/24/cursor-announces-major-update-as-ai-coding-agent-battle-heats-up.html>. Accessed 30 August 2026. Company-provided stat that about 35 percent of Cursor’s own PRs come from cloud agents — not an education finding. [Cited]

CNBC. “OpenAI Rolls Out ‘ChatGPT for Teachers’ for K-12 Educators.” 19 November 2025. <https://www.cnbc.com/2025/11/19/openai-chatgpt-for-teachers-k-12-educators.html>. Accessed 30 August 2026. Not for students; first cohort about 150,000 educators; free through June 2027 **in this report**. OpenAI’s later product page restates free-through June 2028. Do not mash with the AFT Academy 400,000 five-year *target*. [Cited]

CNN Business. “Parents of 16-Year-Old Adam Raine Sue OpenAI, Claiming ChatGPT Advised on His Suicide.” 26 August 2025. <https://www.cnn.com/2025/08/26/tech/openai-chatgpt-teen-suicide-lawsuit>. Accessed 30 August 2026. Allegations. [Cited]

CNA. “All University, Polytechnic and ITE Students to Learn AI Skills from 2027.” <https://www.channelnewsasia.com/singapore/ai-artificial-intelligence-skills-courses-ihl-university-polytechnic-ite-students-2027-6133781>. Accessed 30 August 2026. [Consulted]

CNA. “SUSS Drops AI Detector as More Singapore Universities Question Reliability of Such Tools.” <https://www.channelnewsasia.com/singapore/ai-detection-tool-university-cheating-plagiarism-suss-nus-sutd-sit-smu-6337431>. Accessed 30 August 2026. SUSS disabled Turnitin’s AI writing score; NTU to discontinue from 2027 as “fundamentally unreliable.” [Cited]

CoinDesk. “El Salvador Partners with Elon Musk’s Grok in AI-Powered Education for 1M Students.” 11 December 2025. <https://www.coindesk.com/tech/2025/12/11/el-salvador-partners-with-elon-musk-s-grok-in-ai-powered-education-for-1m-students>. Accessed 30 August 2026. Journalism restating the xAI announcement. Not independent verification. [Cited]

Common Sense Media. *CSM AI Risk Assessment: ChatGPT-5*. 23 October 2025. <https://www.common sense media.org/sites/default/files/featured-content/files/csm-ai-risk-assessment-chatgpt-5-10232025.pdf>. Accessed 30 August 2026. “Teens should not use ChatGPT for mental health or emotional support.” Guardrails weaken in extended conversations. Age 13+ with parental permission; teens bypass by lying about age. [Cited]

Common Sense Media / Youth AI Safety Institute. “Risk Assessments.” <https://institute.common sense media.org/risk-assessments>. Accessed 30 August 2026. Independent third-party ratings. Khanmigo **low risk**, 8 August 2024. ChatGPT-5 **high risk**, 23 October 2025. Character.AI **unacceptable**, 10 April 2025. Social AI companions **unacceptable**, 16 July 2025. Meta AI **unacceptable**, 15 August 2025. Gemini K-12 **high risk**, 9 January 2026. Grok (the **chat** product, not Grok Bot) **unacceptable**, 22 January 2026. Google Search AI Overview / AI Mode **unacceptable**, 14 July 2026. Remaining CSM rating pages other than Khanmigo and ChatGPT-5 were not independently re-opened on the mid2 pass; they match the responsible-use memo. [Cited]

Culatta, Richard. “Profile of an AI-Ready Graduate.” ASCD, 14 July 2025. <https://www.ascd.org/blogs/profile-of-an-ai-ready-graduate>. Accessed 30 August 2026. Headings: Learner, Researcher, Synthesizer, Problem Solver, Connector, Storyteller — the same six as ISTE Live 2026. Do not cite as an Ideator list. [Cited]

EdTech Magazine. “AI Agents for Education: Uses, Examples and Implementation.” December 2025. <https://edtechmagazine.com/higher/article/2025/12/ai-agents-higher-education-transforming-student-services-and-support-perfcon>. Accessed 30 August 2026. Quotes EDUCAUSE Jenay Robert (37 percent chatbot licenses — **chatbots, not agents**) and UT-Knoxville Dan Harder. Johns Hopkins Agent Laboratory (authors claim 84 percent cost decrease) and UT Verse: journalism-sourced; primary pages **not fetched**. [Cited; campus examples UN-VERIFIED against primary docs]

EdTech Magazine. “ISTE Live 26: What Does an AI-Ready Graduate Look Like?” July 2026. <https://edtechmagazine.com/k12/article/2026/07/istelive-26-what-does-ai-ready-graduate-look>. Accessed 30 August 2026. Conference report of the ISTE Profile: Learner, Researcher, Synthesizer, Problem Solver, Connector, Storyteller. Hu-Ann Wren, Patton Elementary (Austin, TX): conference-reported classroom practice, not a districtwide agent program. [Cited]

Fitzpatrick, Dan. “AI Won’t Replace Teachers, but It Will Replace Teachers Who Don’t Use AI.” LinkedIn. <https://www.linkedin.com/pulse/ai-wont-replace-teachers-who-dont-use-dan-fitzpatrick>. Accessed 30 August 2026. Folklore tracking: earliest education-specific use found of the replacement slogan. Not a union, ministry, or peer-reviewed source. Recycled template. [Cited]

Flagler Live. “Artificial Intelligence Integration Rises in Flagler County Schools.” 2026. <https://flaglerlive.com/ai-flagler/>. Accessed 30 August 2026. Grades 6-plus 71 percent; MagicSchool all levels teacher-opened; Khanmigo/Gemini/Brisk 6-12; enterprise ChatGPT; Lightspeed alerts; opt-out; 2,275 PD hours. Implementation facts, not outcome proof. [Cited]

Franczek P.C. “New Artificial Intelligence Developments for Illinois Schools: ISBE Releases Statewide Guidance and Governor Pritzker Signs SB 2909.” JD-Supra. <https://www.jdsupra.com/legalnews/new-artificial-intelligence-9071517/>. Accessed 30 August 2026. Law-firm alert. ISBE primary PDF not fetched; “agentic AI” mention **UNVERIFIED** against the state document. [Cited; ISBE PDF UNVERIFIED]

GlobeNewswire / SchoolAI. “SchoolAI Earns ESSA Tier 3 Certification...” 23 April 2026. <https://www.globenewswire.com/news-release/2026/04/23/3279756/0/en/SchoolAI-Earns-ESSA-Tier-3-Certification-Backed-by-Research-Showing-28-Critical-Thinking-Gain.html>. Pointer to the vendor research page. [Consulted]

Gothamist. “NYC Schools Get AI Guidance Using ‘Red Light, Green Light’ Model.” <https://www.gothamist.com/news/nyc-schools-get-ai-guidance-using-red-light-green-light-model>. Accessed 30 August 2026. [Cited]

Guerra-López, Ingrid. “AI Won’t Replace Teachers—But Teachers Who Use AI Will Change Teaching.” *Education Week*, opinion, 17 October 2025. <https://www.edweek.org/technology/opinion-ai-wont-replace-teachers-but-teachers-who-use-ai-will-change-teaching/2025/10>. Accessed 30 August 2026. Different verb from the folklore slogan (*change teaching*, not *replace those who don’t*). Anti-replacement. [Cited]

Idaho Education News. “School Leaders Grapple with AI Use and Policies for Teachers and Students.” <https://www.idahoednews.org/top-news/school-leaders-grapple-with-ai-use-and-policies-for-teachers-and-students/>. Accessed 30 August 2026. West Ada MagicSchool grades 3-12 teacher rooms; ChatGPT blocked on elementary/middle devices. [Cited]

Idaho Press. “West Ada Adopts Artificial Intelligence Policy.” https://www.idahopress.com/news/local/west-ada-adopts-artificial-intelligence-policy/article_4f57387d-8935-4055-b8e7-a63789ef1f16.html. Accessed 30 August 2026. [Cited]

Palm Coast Observer. “Flagler Schools Stay Ahead of the Curve in AI Education.” 6 May 2026. <https://www.observerlocalnews.com/news/2026/may/06/flagler-schools-stay-ahead-of-the-curve-in-ai-education/>. Accessed 30 August 2026. [Cited]

Reuters. “Google, AI Firm Must Face Lawsuit Filed by a Mother over Suicide of Son, US Court Says.” 21 May 2025. <https://www.reuters.com/sustainability/boards-policy-regulation/google-ai-firm-must-face-lawsuit-filed-by-mother-over-suicide-son-us-court-says-2025-05-21/>. Accessed 30 August 2026. Judge Anne Conway, M.D. Fla., denied early First Amendment dismissal. [Cited]

Singer, Natasha. “Teen Girls Confront an Epidemic of Deepfake Nudes in Schools.” *New York Times*, 8 April 2024. <https://www.nytimes.com/2024/04/08/technology/deepfake-ai-nudes-westfield-high-school.html>. Accessed 30 August 2026. [Cited]

Verma, Pranshu. “A Professor Accused His Class of Using ChatGPT, Putting Diplomas in Jeopardy.” *Washington Post*, 18 May 2023. <https://www.washingtonpost.com/technology/2023/05/18/professor-threatened-fail-class-chatgpt-cheating/>. Accessed 30 August 2026. Texas A&M University-Commerce mass false-positive episode. Cited by NEA 2024. [Cited]

Wonkhe (UK student voice). Quoted in the responsible-use memo: “students’ academic help may be determined by their financial resources rather than their academic competence.” Primary Wonkhe URL was not in the source JSON — **UNVERIFIED**. Wording as the memo recorded it. [Cited via memo; URL UNVERIFIED]

Compiled 30 August 2026. Unfetched = UNVERIFIED. Vendor counts remain labelled as such. Cut claims in the errata are not revived here.