

# The AI Shift in Schools: Risks, Realities, and What's Coming Next

A Goodnotes Research Publication



# Table of Contents

|                                                                                                   |           |
|---------------------------------------------------------------------------------------------------|-----------|
| <b>Abstract</b>                                                                                   | <b>3</b>  |
| <b>Introduction</b>                                                                               | <b>4</b>  |
| <b>Part I: On Mindset</b>                                                                         | <b>8</b>  |
| <b>How can schools build confidence around the use of AI?</b>                                     | <b>8</b>  |
| Toolbox: Student AI Use Survey Template                                                           | 13        |
| Toolbox: Good Future Foundation                                                                   | 16        |
| Toolbox: Foundational AI Literacy Framework                                                       | 18        |
| Toolbox: Anthropic AI Fluency Framework                                                           | 19        |
| <b>Part II: On Mechanism</b>                                                                      | <b>22</b> |
| <b>How can schools support stakeholders to adopt AI responsibly?</b>                              | <b>22</b> |
| Fact Box: Fragmented Regulatory Landscape in the United States                                    | 23        |
| Caution Zone: Unaligned AI Systems and the Failure of Age Verification                            | 25        |
| Fact Box: Implications of EU AI ACT for Education                                                 | 27        |
| Caution Zone: The Hidden Privacy Risks of AI Tools                                                | 30        |
| Fact Box: What is a DPIA?                                                                         | 30        |
| Toolbox: GDPR Requirements for Educators                                                          | 33        |
| Good Practices: What Does Safe Usage Look Like?                                                   | 36        |
| Toolbox: Good Future Foundation Community Platform                                                | 38        |
| Toolbox: AI Risk Assessment (Before Onboarding AI Tools)                                          | 39        |
| Toolbox: DPIA Template (During Onboarding AI Tools)                                               | 40        |
| Toolbox: Example Traffic Light System for Schools (When Rolling Out AI Tools)                     | 44        |
| Toolbox: Good Future AI                                                                           | 47        |
| <b>Part III: On Motivation</b>                                                                    | <b>50</b> |
| <b>How can schools inspire stakeholders to change their behaviors towards new best practices?</b> | <b>50</b> |
| Good Practices: Assigning AI Pedagogical Roles to Support Learning                                | 59        |
| Good Practices: Designing AI Prompts to Support Learner Agency                                    | 61        |
| <b>Conclusion</b>                                                                                 | <b>64</b> |
| <b>Acknowledgements</b>                                                                           | <b>66</b> |
| <b>References</b>                                                                                 | <b>67</b> |

# Abstract

Generative AI is accelerating change in schools. It is lowering the cost of information and knowledge while unsettling assessment, authorship, and trust. This paper reports on research that combines educator interviews, case studies from early adopters, and a review of recent scholarship to map the opportunities and the risks. The analysis identifies three consistent and relevant challenges: how to build confidence in safe and effective use, how to support educators and learners in adopting AI responsibly, and how to sustain motivation for change without losing sight of educational purpose. Findings show that compliance and policies by themselves do not address risks of bias or privacy, and that a broad range of stakeholders in schools are seeking both conceptual clarity and practical strategies. We propose a framework of Mindset, Mechanism, and Motivation as a way to guide responsible adoption. This model recognizes that AI can and should be responsibly integrated into school life as a form of support and guidance, but within the values and practices that shape teaching and learning.

# Introduction

The use of Generative Artificial Intelligence (GenAI) has become increasingly widespread across all areas of life, learning, and work. This paper serves as an update concerning the current state of play and the implications for schools when it comes to implementing responsible use and best practices. We know that GenAI is here to stay and that it is already altering the way that teachers teach, learners learn, and senior leaders lead their schools. In this research, we examine important areas that have gained prominence in current educational discourse and require urgent attention from educators and policymakers, including safeguarding and protection around AI use, human-AI collaboration in academic work, and fundamental questions surrounding learning and education in the age of AI.

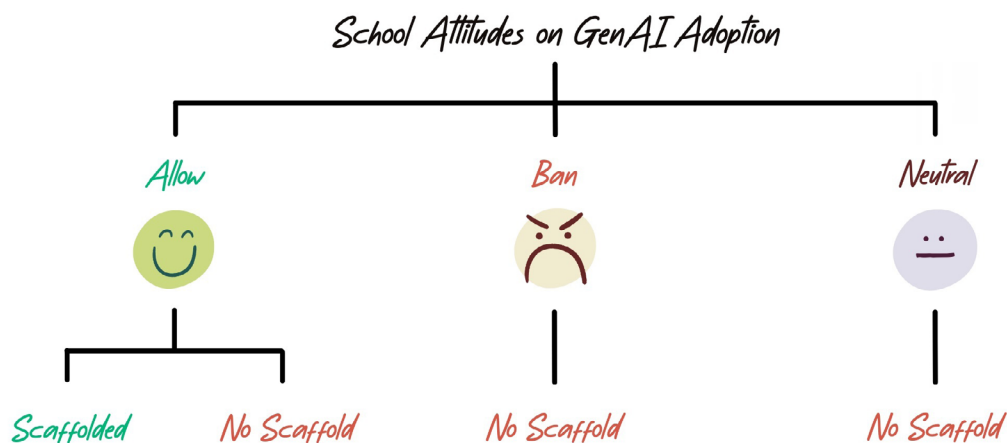
While our earlier publications identified many of the key challenges and emerging practices around AI in education, the pace of change has accelerated beyond what many had expected. Indeed, critical voices have emphasized not to put technology front and center, but instead focus on the interventions and their impact on learners and educators within broader frameworks of ethics, values, power dynamics, and interdisciplinary approaches; however, there remains limited scaffolding on how educational communities can actually shape responsible AI implementation in practice (Holmes et al., 2025).

Therefore, in our conversations here with a diverse range of stakeholders, we seek to understand and document how schools are experiencing and adopting AI technologies.

We acknowledge that there are many concerns around AI in education, which we discuss throughout the paper. However, we also emphasize that there are undeniable benefits when AI is thoughtfully implemented. **The current stake, though, is that we are not merely undergoing a technological transition, but rather a paradigm shift on educational purpose and agency, with pedagogical implications on how teaching and learning happen, who controls the design and implementation of technologies in education processes, and what it means to reform coursework and assessments to evaluate genuine understanding and produce meaningful work both within and beyond the classroom.**

Rather than taking an instrumentalist view of AI as a set of tools to be deployed, our focus here is on how schools' approaches to adoption shape learners' growth and wellbeing. In our previous research, we argued that while schools may have varying degrees of risk tolerance, concerns, and openness to innovation, it is those who actively engage in multistakeholder conversations and development that are best positioned to thrive:

When it comes to schools formally adopting AI, we distinguish them by three categories: those that allow AI, those that ban AI, and those with a neutral or uncertain stance. Inevitably, there is no scaffolding provided by schools that take on the latter two stances. However, even for schools that allow students to use AI, they might not have developed a systematic approach that directs students to understand how, why, and under what circumstances AI could be used to enhance their learning experiences (Goodnotes, 2024).



Enabling education stakeholders with active and scaffolded engagement, however, is not without its challenges. **We recognize that AI in education is fundamentally a balancing act between concerns and benefits, and while navigating this path requires difficult work, we believe there is much we can gain if we approach it thoughtfully.** As such, in this paper, we present the current challenges and paths forward with regard to the following three guiding questions:

- **On Mindset:** How can schools build confidence for effective use of AI?

- **On Mechanism:** How can schools support stakeholders to adopt AI responsibly?
- **On Motivation:** How can schools inspire stakeholders to change their behaviors towards new best practices?

Why do these questions matter? We focus on two aspects that remain integral to school life: learning and safeguarding.

On the aspect of learning, AI is driving the reduction of the cost of intelligence, meaning tasks that previously required expensive human cognitive labor can now be done cheaply and quickly by AI systems. However, **unlike other industries where AI can efficiently automate certain tasks and services, education has a distinct problem—that intelligence is both the system’s input and output**, that assessment is built around intelligence as both what learners bring in (their cognitive abilities, prior knowledge, etc.) and what they are expected to develop (becoming more capable and intelligent) in order to produce high-quality work. Education is therefore uniquely vulnerable, because if the ‘products’ of intelligence (well-written essays, coherent arguments, solved problems) can be generated effortlessly, short-circuiting the entire cycle of learning, how do we then distinguish between genuine understanding and growth versus performance where ‘AI handled the thinking parts’?

When it comes to safeguarding, AI is increasingly ubiquitous across educators’ and learners’ academic and social lives. With AI now integrated into search results and daily platforms, it does not take very much at all for one to engage with these tools. However, more work needs to be done to encourage responsible use among both educators and learners. What are the implications when personal information is input into certain models? How do we know which platforms are

appropriate for different types of data sharing? Educators and learners need to understand the implications of sharing data and personal information with AI systems, and recognize that AI-generated content in many cases can be inaccurate, misleading, or fabricated.

In short, **now that people are using AI, how do we make sure that this will make things better, not worse?** We believe that we must build confidence, provide support, and motivate education stakeholders to cultivate critical understanding and effective use of AI. Of course, similar to how we have noted not to conflate ‘blind trust’ (denoting uncritical adoption and dependence on AI) with ‘demystified trust’ (denoting scaffolded adoption with a critical understanding of AI) in our previous publication, we emphasize that there are multiple nuanced layers to how we define and approach confidence, support, and motivation. For instance, some recent research has identified that higher AI literacy, confidence, and trust in AI may be associated with reduced critical engagement, whereas confidence in one’s own capabilities supports greater critical thinking (Wijaya et al., 2024; Lee et al., 2025). **Does that mean we should ban the use of AI?** We argue that these results in fact increase the urgency for establishing common grounds for articulating what critical AI literacy and confidence mean in the first place, so that the subsequent support mechanisms and motivation are aligned.

**When knowledge and understanding of the implications of AI remain fragmented across our educational communities, the entire system is less responsive, limiting our collective ability to safeguard students and preserve the transformative human connections that make learning meaningful.** It is with these considerations in mind that we turn to examine the current landscape through multistakeholder perspectives. In the following sections, we synthesize insights from relevant literature alongside educator and learner voices to explore practical pathways for building AI confidence, supporting responsible adoption, and motivating behavioral change toward best practices.

Throughout this analysis, we also provide practical guidance through ‘Good Practices’ boxes that highlight effective strategies, ‘Caution Zones’ that warn against critical risks and vulnerabilities, ‘Fact Boxes’ that clarify key regulatory and technical concepts, and ‘Toolboxes’ that offer resources and useful frameworks that any schools and educators can leverage and adapt for use. Our commitment is to a future where learning is not reduced to what can be easily automated and measured, but rather amplified to sustain our deepest educational values—the cultivation of wisdom, empathy, and most importantly, our shared capacity for meaning-making and flourishing.

## Part I: On Mindset

### How can schools build confidence around the use of AI?

The promise of AI in schools is met with an equal measure of apprehension. Before any policy can be written or any tool can be deployed, the primary battle must be fought on the terrain of mindset. It is here, in the attitudes and beliefs of educators and learners, that the foundations for either responsible innovation or reactive fear are laid. This section addresses this first and foremost challenge: How do we transition our school communities from a place of fear and mistrust to one of critical confidence and constructive engagement?

#### **Current Challenges: Fear, Misunderstanding, and the Honesty Gap**

**Ubiquitous Infrastructure, Ubiquitous Adoption, and Uncertainty**  
One of the primary reasons educators lack confidence around AI is that the technology has evolved from an optional tool to an invisible infrastructure. This creates anxiety about what learners are actually doing and whether traditional guidance on AI tool use is enough.

Across our conversations, we noticed that educators are beginning to describe a landscape where AI has become infrastructural—integrated so deeply within daily practices that its use often occurs without explicit recognition. Rachel Bowen, Assistant Head at Headington Rye School in Oxford, stated that AI has

“become a lot more embedded into things that we’re using... Rather than just using things like Gemini, Google does that little summary at the top.” Similarly, Dr Anne-Marie Stanton-Ife, Deputy Head Academic, also at Headington Rye, noted, “Even the plagiarism check, sample check on Google Classroom... I think a lot of people don’t realize they’re using it because it’s sort of there by stealth.”

**Such ‘stealth’ quality by design highlights how AI is becoming increasingly ubiquitous—it is now part of a pervasive background infrastructure that subtly influences human decision-making by pushing new content (Du, 2024).** This alters, if not reverses, the agentic relationship between humans and AI, creating uncertainty for educators: **How can they guide student learning when they themselves do not have confidence in their understanding of the AI tools that now mediate that learning?**

The challenge for building educator confidence is compounded by the reality that learners are driving AI adoption from the ground up, often without institutional permission or guidance. Selin O., a final year student at Robert College in Istanbul, described the widespread use: “I don’t know a single person who would look up the definition on Google and then go back to that paper and write it. Everyone would just put it on ChatGPT and get the

answers.” She further reflects on how her peers’ gravitation toward AI is not just about convenience, but also their comfort with technology over authority figures like teachers, “Maybe it’s because your computer or your phone is the thing that is closest to you, and you know that its probability of judging you is way lower than the probability of your teacher judging you.” In other words, learners may turn to AI when they feel the emotional need for non-judgmental support. This also highlights the relational aspect of learning that effective AI integration needs to take into consideration.

Furthermore, when asked about whether rules for AI use are established during group work, Conrado Torres, a student at Torcuato Di Tella University, responded stating AI use in collaborative projects often emerges organically, “There weren’t any boundaries – we didn’t sit and say ‘we are going to use this tool’ or ‘we are not going to use anything at all.’” Instead, his peers naturally integrated AI as part of their collaborative project workflow. As such, unlike other classroom technologies that are much easier to regulate, the organic AI adoption by learners leaves educators feeling much more uncertain and unprepared: Students may be more fluent with AI tools than their teachers, traditional methods of detecting or understanding student work may no longer apply, and the fundamental assumptions about independent learning are being challenged daily.

The implications of this ubiquity shift thus extend far beyond technology integration.

**If AI use becomes as automatic and habitual as reaching for a calculator, opening up a web browser, or even doomscrolling through social media—to the point that using AI is no longer a deliberate action in itself but rather an accepted presence that shapes how we interact, then schools must reconsider their fundamental assumptions about what constitutes independent work, how to assess genuine understanding, and what skills students need to develop for their futures.**

The question, then, is no longer whether AI will (or should) be integrated into school life, but **how schools can help educators develop the confidence to maintain educational purpose and intentionality in an environment where AI capabilities are simply an inevitable part of the background infrastructure of learning** (Gulson et al., 2022).

### **Secret Cyborgs & Double Standards**

Another barrier to building institutional confidence around AI is the hidden use of AI, or what Mollick (2025) calls the ‘secret cyborg’ problem—the discrepancy between the reported and actual use of AI. In the context of education, the ‘secret cyborg’ problem exists on both the learners’ and educators’ ends. This creates an honesty gap where neither party knows what the other is actually doing with AI, undermining confidence and trust on both sides. First, on the learners’ end,

many seem to be reluctant to admit their use of AI when approaching homework and coursework. Steph Chambers, Digital Innovation Lead at ACS Egham, surveyed her students about their AI use, and explained that despite almost all of her students using AI, “they are very hesitant to admit using it.” She further elaborated that there is a “preconceived notion among students that teachers don’t like it or they don’t like students using it,” on the other hand, “when you ask teachers, they say they just don’t understand it and they’re scared.”

The reluctance to acknowledge AI use despite its prevalence points toward a misalignment between student experience and institutional or educators’ expectations, where AI use is by default stigmatized as cheating rather than recognized as a potential learning tool (Bao & Zeng, 2025). As such, the implications of this miscommunication extend far beyond simple policy compliance: **When learners hide their AI use, they lose confidence in seeking appropriate guidance, while educators lose confidence in validly assessing learning and providing meaningful feedback.** Schools then lose the opportunity to guide and scaffold that use constructively or to understand how AI is actually shaping student learning.

In fact, this pressure is even more acutely felt in collaborative settings. Naomi King, a Design and Computer Science major at Queensland University of Technology and member of the Goodnotes Customer Support team, described the tension when working with peers who rely heavily on AI tools such as Cursor to complete their assignments:

**It’s caused a bit of tension, to be honest, because you can’t accuse them of using AI if you don’t have hard evidence, but also it’s likely very clearly, there’s no way they could perform at that level... Especially in group work, you can see other teammates doing it, and they are way more productive than you, and it looks like they’re carrying the assignment. It’s really hard to not use AI just so you can match them.**

This highlights how the lack of transparency actively pressures conscientious and responsible students to compromise their own learning process to keep pace.

From the perspective of those working in education, the honesty gap is exacerbated by what Tal Havivi, Managing Director for Research & Development at ISTE, identified as a ‘double standards’ problem:

**And then even within that, there are questions around double standards, right? If educators are allowed to use certain AI tools to develop lesson plans and help them with their work, but students are held to a different standard for the work that they can generate.**

In other words, when learners observe educators using AI while being prohibited from doing so themselves, this inconsistency undermines both institutional policy credibility and reinforces secretive AI use, reinforcing the cycle where neither party has confidence in the system’s fairness or effectiveness.

## **Paths Forward: Transparency, Literacy, and Shared Principles**

The challenges outlined above—ubiquitous AI infrastructure, clandestine AI usage, and institutional uncertainty—may seem overwhelming, but they point toward a clear first principle: transparency. Rather than attempting to control or detect AI use through technological means, we need to create the conditions for honest dialogue and thoughtful experimentation. Yet transparency alone is insufficient without the literacy to understand the implications of AI use, and the shared principles to guide how we act on that understanding. In the following, we elaborate on how confidence—whether in educators’ ability to guide AI use, learners’ capacity to use it responsibly, or institutions’ capability to support both—cannot be built on surveillance or suspicion, but must emerge from authentic relationships grounded in mutual understanding and shared purpose.

### **Transparency as Trust-Building**

The challenges outlined above—ubiquitous AI infrastructure, secretive AI usage, and institutional uncertainty—may seem overwhelming, but they point toward transparency as an essential principle. As implied from discussions above, the challenge schools face in building institution-wide confidence in the age of AI is not just technological but also relational: **How do we create authentic relationships between students and teachers when traditional assumptions about academic work no longer hold?**

Across our conversations, we found that some educators have begun experimenting with encouraging

complete transparency, recognizing that mutual disclosure of AI use directly addresses the fear and misunderstanding that undermines confidence. When Steph Chambers surveyed her students about their AI use, it wasn’t to “catch anyone out” but to “learn about the different trends and perspectives” to better support her students in navigating AI’s potential and challenges (see Toolbox: Student AI Use Survey Template).

Likewise, Kate Atherton, Associate Assistant Headteacher at St John the Baptist School in Surrey, deliberately encourages mutual disclosure of AI use: “It’s been nice having done the talks. Students have come to me and be like, ‘Miss, so do you use AI to help with your lesson planning?’ I was like, ‘Yes, I did.’” Upon acknowledging her own AI use, she explained her reasoning to the students. She quickly observed that students began approaching her to openly discuss their experiences, sharing how they used AI for specific subjects and seeking guidance on appropriate boundaries. **As demonstrated, the transparency initiated on the educators’ end created the conditions for reciprocal honesty, transforming the classroom dynamic from one of surveillance and suspicion to one of collaboration and learning.** This approach is further substantiated by Dr Maria Ruiz-Primo, Associate Professor at Stanford Graduate School of Education:

**I do think that we should encourage both sides. Just as I encourage students to be honest when they use ChatGPT and how and for what, I think that I should also let them know if I am using an AI tool.**

In other words, this mutual transparency creates psychological safety for genuine conversation about appropriate and inappropriate AI use.

However, building trust through transparency cannot remain at an individual voluntary level; rather, it must be institutionalized through aligning individual practices with broader professional standards and organizational expectations. **Much of the lack of confidence around AI use does not stem from technical incompetence, but rather from operating in grey zones without clear ethical standards. When neither educators nor learners have clarity on whether their practices are appropriate, this uncertainty creates guilt, encourages fear-based secrecy, and undermines confidence.** In this sense, institution-wide procedural transparency is not just compliance for compliance's sake; instead, it itself is a mechanism for sustained trust-building that enables confidence.

Some institutions are already putting such policies into practice, shifting the focus from policing to process. Naomi King described the straightforward requirement in her design program:

**You have to commit to AI transparency at the start of every assignment where you specify what you used it for... I would always state what it was and then, in the appendix, show the prompts I used and the outputs I received. That's the policy for design school.**

This is simple, but mandating such disclosure encourages students to think critically about how they use AI as a tool rather than hiding its use altogether. This approach also helps make staff more aware of their own responsibilities. As Kate Atherton noted:

**I think it's also making staff more aware of what they need to ask consent for from the students. So actually if they're going to put it in something like Grammarly to check, they shouldn't be doing that without talking and asking the students about it.**

Rather than treating transparency as monitoring whether students used AI, it should be sought as a pathway to more authentic relationships and meaningful dialogue about learning, a pedagogical opportunity that builds both learners' confidence in their learning and educators' confidence in guiding that learning effectively.

# Toolbox: Student AI Use Survey Template

(adapted from Steph Chambers)

This survey works best when administered early in the academic year or before introducing new AI policies. It can be done anonymously to encourage honest responses. Use results to understand current usage patterns within the class and tailor your approach to AI education. Section 2 tells you about adoption rates; Section 3 reveals how students are actually using AI; Section 4 helps identify knowledge gaps that need addressing; and Sections 5–6 reveal the disconnect between student perceptions and school practices.

Hi students,

I'm gathering your thoughts about how you use Artificial Intelligence (AI) in your learning. This survey helps me understand current trends and better support you in navigating AI's potential and challenges.

Please note: This is not about catching anyone doing something wrong. I want to understand how AI is being used so I can adapt my teaching to best serve your learning needs. All responses are confidential and will only be used to improve our learning environment. Your honesty is appreciated!

## Section 1: Basic Information

Grade/Year:

Subject/Class:

## Section 2: AI Usage

### 1. Have you ever used AI tools for schoolwork?

Yes

No

Not sure

If you answered "No" or "Not sure," you can skip to Section 4.

## Section 3: AI Experience (For AI Users)

### 2. Which AI tools have you used? (Check all that apply)

ChatGPT

Microsoft Copilot

Google Gemini

Anthropic Claude

Grammarly

Canva AI

Perplexity AI

QuillBot

Other:

**3. What do you use AI for in your schoolwork? (Check all that apply)**

Research and finding information  
Writing assistance (grammar, style, structure)  
Brainstorming and generating ideas  
Studying and summarizing information  
Creating visuals or presentations  
Translation  
Understanding difficult concepts  
Checking my work  
Other:

**4. In which subjects have you used AI? (Check all that apply)**

English/Language Arts  
Mathematics  
Science  
Social Studies/History  
Foreign Languages  
Art/Creative subjects  
Computer Science  
Other:

**5. Describe one time you used AI successfully in your schoolwork.**

Example: helping with brainstorming, explaining a concept, creating something

**Section 4: AI Understanding and Ethics**

**6. Are you aware of any concerns or ethical considerations when using AI for schoolwork?**

Yes  
Somewhat aware  
No

**7. If you use AI for writing or research, how do you avoid plagiarism? (Check all that apply)**

I cite when I use AI  
I use AI only for ideas, then write myself  
I edit and revise AI output significantly  
I check my work with my teacher  
I'm not sure how to avoid plagiarism  
This doesn't apply to me

**8. How do you check if information from AI is accurate?**

I verify with other sources  
I use multiple AI tools to compare  
I trust the AI is correct  
I ask my teacher  
I don't usually check  
This doesn't apply to me

## Section 5: Teacher and School Perspective

### 9. Have your teachers discussed AI use in class?

- ☐ Yes, regularly
- ☐ Sometimes
- ☐ Rarely
- ☐ Never

### 10. What is your teachers' general attitude toward AI use? (As you perceive it)

- ☐ Very supportive
- ☐ Somewhat supportive
- ☐ Neutral/unclear
- ☐ Somewhat concerned
- ☐ Very concerned
- ☐ I'm not sure

### 11. When do you think students should NOT use AI for schoolwork?

## Section 6: Impact and Future

### 12. How has AI impacted your learning? (Check all that apply)

- ☐ Made research faster and easier
- ☐ Improved my writing
- ☐ Helped me understand difficult topics
- ☐ Made me more creative
- ☐ Made me rely too much on technology
- ☐ Reduced my critical thinking
- ☐ Caused concerns about academic honesty
- ☐ No significant impact
- ☐ Other:

### 13. Do you have concerns about AI or the future of AI?

- ☐ No concerns
- ☐ Minor concerns
- ☐ Some concerns
- ☐ Major concerns

### 14. What concerns (if any) do you have about AI?

## Section 7: Feedback and Suggestions

### 15. What can our school/teachers do to help you better understand and use AI appropriately?

### 16. If you could tell your teachers one thing about AI use, what would it be?

### 17. Any other comments or questions about using AI in your education?

### **Cultivating Foundational AI Literacy & AI Fluency**

While transparency creates the conditions for trust, it is insufficient on its own. For educators and learners to engage with AI confidently, they must possess foundational AI literacy to critically understand the potential, limitations, and ethical implications of the technology itself. As Tal Havivi framed it, the focus on AI literacy is “Not so that every teacher is an AI expert, but so every teacher has a strong enough understanding of how this technology works, how it can work for them, and things to be mindful of and to stay away from.”

This call for a ‘strong enough understanding’ of AI is not merely anecdotal, but substantiated by large-scale research. An international study by Viberg et al. (2024) surveying over 500 K-12 teachers across six countries found that an educator’s confidence is directly tied to their AI self-efficacy and fundamental understanding of the technology. The research-

ers demonstrated that when teachers have a better grasp of AI, they perceive more benefits and fewer concerns, which in turn builds their trust. These perceptions are powerful, as perceived benefits and concerns alone explained 49% of the variance in teacher trust. Furthermore, while demographic factors like gender or age did not significantly impact trust, cultural and geographic differences did. For instance, teachers with higher uncertainty avoidance and a stronger long-term orientation reported higher trust. Geographically, teachers in Norway, Sweden, and the U.S. perceived fewer benefits and more concerns than their counterparts in Brazil, Israel, and Japan. This underscores the need to invest in professional development programs that take culturally sensitive approaches to improve teachers’ realistic understanding of AI and build their self-efficacy.

## **Toolbox: Good Future Foundation**

It is worth noting here that Good Future Foundation, a UK registered charity funded by Goodnotes’ CEO and Founder, Steven Chan, offers tech agnostic support for schools across the UK and beyond. The organization focuses specifically on responsible use and strategic implementation of the technology through their AI Quality Mark and in person professional development.

For more info, please visit: <https://www.goodfuture.foundation/>

What, then, constitutes this 'strong enough understanding'? It involves several layers. First, AI literacy requires looking under the hood to grasp AI's architecture and the data origins. As John Jones, Assistant Head at Royal Grammar Schools Worcester, explained:

**When you work out how they work, you kind of automatically start to think, well, there are issues there, there are concerns there, because there's training data, who's got access to it, who's behind it, who's trained it in the first place.**

Second, AI literacy demands a critical awareness of the tool's inherent fallibility. Nathan Nagaiah, Lead of UK Centre for AI and Strategic Lead for Artificial Intelligence, Data and Partnerships at London Borough of Newham, stressed this point, warning that AI "can easily hallucinate itself, so it might not even give you the true information. So it's really important that you know, yes, we have to embrace AI, but at the same time also take into consideration the high risk that is associated with it."

Finally, AI literacy reframes the human's role as the indispensable expert. The tool amplifies, but does not replace, human knowledge. As Ha Thi Hai Nguyen, Dr of Education Candidate and Lecturer at Vietnam National University in Hanoi, observed from her own practice, to get quality results, "You are the one who really needs to master the rules well or the frameworks well." In other words, people who can effectively and confidently leverage AI are those who already have deep knowledge of their domain and can clearly articulate their intentions.

There are, of course, many AI literacy frameworks and programs that schools are adopting, including but not limited to AI4K12 (AI4K12.org, 2021), MIT's The DAILY Curriculum for Middle School Students (Breazeal & Lee, 2022), and UNESCO AI Competency Framework for Teachers (Miao & Cukurova, 2024). Below, we have adapted a foundational AI literacy framework by Chan and Colloton (2024) for the K-12 context for your reference.

# Toolbox: Foundational AI Literacy Framework

## (Chan & Colloton, 2024) – Adapted for K-12

### AI Literacy Dimension

### Description for K-12 Context

#### AI Concepts

Familiarity with essential AI terminology and concepts (such as artificial intelligence, narrow/general/super intelligence, machine learning, and hallucinations) to support understanding of how AI systems function in both teaching and learning contexts.

#### AI Applications

Awareness of common AI tools and applications used in educational settings and everyday life, including learning platforms, virtual assistants, recommendation systems, and image recognition technologies.

#### AI Affectiveness for Human Emotions

Understanding how AI systems can detect and mimic human emotions, including the implications and considerations for using affective AI in classrooms, student support, and personal interactions.

#### AI Safety and Security

Awareness of potential security risks and challenges associated with AI applications in educational environments, including concerns about student data privacy, personal information protection, and safe technology practices.

#### Responsible AI Usage

Developing responsible practices when using AI tools for teaching and learning, understanding that AI systems have limitations and may produce incorrect information, considering ethical implications, and maintaining critical evaluation of AI-generated content and results.

**Educators should recognize that cultivating AI literacy requires a highly proactive process involving continuous use and thoughtful experimentation.** Hence, we need to not just be literate but also fluent: In the study of languages, literacy generally refers to being able to write and understand a language, while fluency goes beyond literacy to enable one to create new things from the language to express themselves (Campbell et al., 2023). This can be extended to the context of AI: While AI literacy refers to the foundational knowledge that enables one to safely and critically understand, use, and evaluate AI

(Chan & Colloton, 2024), AI fluency builds on this foundation to empower one to appropriately and effectively leverage AI to creatively reimagine processes and solve problems innovatively (Bernhardt, 2025).

Below, we also provide Anthropic's AI fluency framework, developed by Dakan and Feller (2025), where they outlined 'delegation', 'description', 'discernment', and 'diligence' as four interconnected competencies and steps necessary for effective interactions with AI.

## Toolbox: Anthropic AI Fluency Framework (Dakan & Feller, 2025)

| Competency  | Definition                                                                                                                                                                                                                               |
|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Delegation  | Setting goals and deciding whether, when, and how to engage with AI (i.e., making thoughtful decisions about what work is appropriate for you to do, for AI to do, or for you and AI to do together, and how to distribute those tasks). |
| Description | Effectively describing goals to prompt useful AI behaviors and outputs (i.e., communicating with AI in ways that create a productive collaborative environment).                                                                         |
| Discernment | Accurately assessing the usefulness of AI outputs and behaviours (i.e., thoughtfully and critically evaluating what AI produces, how it produces it, and how it behaves).                                                                |
| Diligence   | Taking responsibility for what we do with AI and how we do it (i.e., being thoughtful about which AI systems you use, being honest about the role of AI in your work, and taking responsibility to verify the outputs you use or share). |

It is through hands-on engagement coupled with effective professional development that educators and learners become critical users who not only understand how to responsibly and effectively use AI tools, but also remain cognizant of their capabilities and constraints, so that they can leverage AI strategically as an enabler rather than a crutch.

### Putting Visionary Leadership into Action

With transparency, literacy, and fluency laying the groundwork for individual confidence, a further step in building institutional confidence is to structure that potential into directed action. As Mollick (2025) suggests, individual performance gains with AI do not automatically translate to organizational improvement.

The first act of leadership is to move beyond simple declarations of urgency

and instead paint a clear picture. This vision must answer the crucial, unspoken questions that fuel educator and learner uncertainties: **What will our work and learning look like? How do we bring that in line with curriculum expectations? How will we be supported and rewarded for engaging with these new tools?** It is this vision that provides the context and purpose for any formal principles that follow. The process of creating these principles also becomes an act of communal buy-in rather than top-down compliance. As Dr Anne-Marie Stanton-Ife implied, this often begins with collaborative conversations among “heads of department sharing ideas and research” before being formalized into a clear and accessible consensus, as John Jones described, “so every single child, every single member of staff has to sign an acceptable use policy before they can use artificial intelligence on the network.”

**“But urgency alone isn’t enough. These messages do a good job signaling the ‘why now’ but stop short of painting that crucial, vivid picture: what does the AI-powered future actually look and feel like for your organization?”**

**– Professor Ethan Mollick, 2025**

**Associate Professor at the Wharton School of the University of Pennsylvania**

Yet, even the most thoughtfully crafted principles remain inert on paper. They must be animated through active and authentic leadership that models the desired behaviors, **establishing clear incentives that champion, rather than punish, experimentation and the sharing of new discoveries.** As suggested by Dr Guillermo Solano-Flores, Professor at the Stanford Graduate School of Education, the most constructive dynamic is when leaders position themselves as “co-learners” along-

side the learners and colleagues, to foster a culture where asking questions is more valuable than having all the answers, and to give the learning community a direction and a safety net to confidently move forward with purpose.

Finally, beyond principles and behaviors, this leadership modeling must extend to cultivate what Tal Havivi described as authentic digital citizenship:

**So not just ‘here are all the things that you should not do—don’t do this, don’t do this, stay away from this.’ But more so, like, ‘what does that actually mean to be a productive citizen in a digital world?’**

This approach shifts the discourse towards capability building, supporting both learners and educators to continuously update their critical awareness under a rapidly evolving digital landscape and to systematically reflect on the ethical and personal approaches to digitalization in education (Raffaghelli & Stewart, 2020). As John Jones articulated, “We want children

to be able to regulate their own use and ask critical questions about the consumption that they are experiencing,” which is essential “because [educators] can only protect children up to a certain point within the boundaries of our school and our networks.”

Through establishing shared principles through collaborative vision-setting, leading by example, and a focus on developing digital citizenship rather than mere compliance, schools could cultivate the collective confidence necessary for meaningful AI integration.

## Part II: On Mechanism

### How can schools support stakeholders to adopt AI responsibly?

While a confident mindset is necessary, it itself is insufficient for responsible AI adoption. As such, the challenge shifts from the psychological to the practical: building the mechanisms of support. Yet, schools find themselves caught between the demand for robust safeguarding and the practical impossibility of vetting every tool, creating a chasm between compliance on paper and safety in practice. This section therefore investigates a critical operational question: How can schools forge practical, effective support systems in a landscape of overwhelming complexity and constant change?

#### Current Challenges: Compliance & Awareness Gaps

**External Policies: Compliance Requirements as an Illusion of Safety**  
When it comes to implementing AI tools in educational settings, one of the largest concerns is around data privacy and protection. Because schools cannot host their own AI systems the same way they curated and controlled resources like before (e.g., their own libraries, textbooks, record-keeping systems), they are compelled to adopt external systems as part of their everyday workflow. More concerningly, LLMs themselves introduce new threat vectors, such as being able to infer and identify personal attributes from texts even when Personally Identifiable Information (PII) is explicitly anonymized

(Stabb et al., 2024). All of these factors dramatically increase both the potential consequences of data misuse and the need for robust due diligence to a level that many schools are not accustomed to.

For the past two years, the regulatory environment governing AI in education has been undergoing rapid changes, accompanied by compliance obligations at multiple levels and aspects. **However, such regulatory obligations are often vague, overwhelming, and delayed, creating an illusion of safety through requiring massive administrative paperwork, while consuming the resources that could actually protect students.**

For instance, at the international level, the UNESCO AI and education: Guidance for policy-makers have provided broad guidance with “overarching principle for AI and education policies” to address critical societal issues of “fairness, transparency, accountability, human rights, democratic values, bias, and privacy” raised by AI (UNESCO, 2021, p. 32), advocating for “establish[ing] data protection laws which make educational data collection and analysis visible, traceable, and auditable by teachers, students and parents” (p. 33). This broad statement, however, may be toothless in practice, as it creates for teachers and school administrators vast amounts of paperwork with technical and nuanced

details that rapidly evolve to keep pace with technological change. Similarly, the United States presents a fragmented patchwork of different state policies alongside inadequate federal legislation that leaves wide gaps in protection (see Fact Box: Fragmented Regulatory Landscape in the United States), while the

EU AI Act introduces legally enforceable compliance mechanisms but creates extensive administrative burdens on top of technical impossibilities. These regulatory approaches, whether insufficiently comprehensive or overly prescriptive, fail to provide schools with practical guidance for improving actual safety outcomes.

## Fact Box: Fragmented Regulatory Landscape in the United States

In the United States, the regulatory landscape for AI in education is still fragmented. While the federal government has set up the Artificial Intelligence Education Task Force to “promoting the appropriate integration of AI into education” (The White House), what that means in practice remains unclear.

As of July 2025, 26 states have issued their own official guidance or policies on the use of AI in K-12 schools (AI for Education, 2025). While these efforts are commendable, this inevitably creates a patchwork of aspirational yet ununified approaches that make it untenable for edtech providers and AI companies operating across state lines.

While there are relevant legislations on a federal level, they remain insufficient. For example, FERPA (Family Educational Rights and Privacy Act of 1974) gives parents control over children’s educational records with rights transferring at age 18 or upon postsecondary enrollment; however, it has been largely criticized for its practical lack of protection, especially when it comes to broader educational risks of AI beyond privacy and data input.

Meanwhile, COPPA (Children’s Online Privacy Protection Act of 1998), which requires parental consent for online data collection from children under 13, not only leaves out a large protection gap for young adults aged 13-17 by treating them as capable of self-consent, but also creates confusing inconsistencies with FERPA’s age definitions (Do, 2025). Notably, COPPA has been recently amended to include broadened definitions of personal information including biometric identifiers, stronger parental consent for data collection, and stricter security requirements (Marlowe, 2025).

Nonetheless, both FERPA and COPPA still remain as narrow privacy-centric approaches, which are far from being comprehensive enough to address the full spectrum of concerns associated with critical developmental, pedagogical, and accessibility risks AI poses in education beyond data concerns (Do, 2025).

Meanwhile, the absence of timely and practical guidance from educational authorities forces schools to develop their own support frameworks without institutional backing or shared best practices. As Dr Anne-Marie Stanton-Ife pointed out, “It wasn’t helpful that the DfE (Department for Education) took so long to come up with that advice [on AI use in schools]. You know, it should have come out a year ago, really, at least.” This pattern of policy lag continues to create for schools an ‘impossible situation’. For instance, in the UK, the DfE’s Keeping Children Safe in Education (KCSIE) 2025 guidance—which promised a substantive update—was not published fully until September 1st. As a result, many designated safeguarding leads are approaching the new academic year without the framework needed to update policies and train staff (Norden, 2025).

This regulatory vacuum is particularly concerning given the accessibility of AI: Unlike previous institutional technology procurement, AI adoption can easily happen at the grassroots level, regardless of top-down policies or institutional readiness. Learners and educators are already using AI tools without adequate safeguarding protocols, while tech companies sell solutions that schools may be unprepared or have the capacity to evaluate safely.

Even when guidelines do exist, their update cycles cannot keep pace with AI development. For example, Steph Chambers, revealed that IB updates its guidance around AI use only “every year.” While having annual revisions is commendable, the very length of the cycle results in a built-in lag from top institutional bodies, as Chambers further described the challenge of aligning school policies with IB’s requirements:

**There was a group of eight or nine [teachers] at the school that put together that piece [of academic integrity policy] along with the AI piece, as well as trying to align it with what the IB states, because we are responsible to them. But [IB] up until now has been quite vague about what they say is acceptable.**

The cumulative effect of delays and vagueness trickling down from top authorities forces schools into a reactive position and an impossible situation: **They must simultaneously interpret unclear guidance, maintain accountability to such external standards, and take initiative to develop internal frameworks without adequate institutional backing.** As such, schools are left with no choice but to engage in superficial compliance busywork, while knowing that systemic issues like outdated curricula, slow-to-react responses, and lack of up-to-date training remain unaddressed by the very authorities creating these administrative burdens.

However, this is not to say that such administrative tasks cannot be made more efficient. For instance, although a Data Protection Impact Assessment (DPIA) cannot be fully automated, many of its procedural components can be. A school can proactively design a streamlined system to manage these assessments, using tools to automatically flag when a DPIA is needed, collate system details, and suggest risks and controls. The key is keeping a human in the loop for tricky judgment calls, but letting automation handle the repetitive parts to make the process faster and more consistent.

We must recognize that safety in education is not something legislation alone can ensure—laws and compliance frameworks account for only a small

subset of the actual safety work, while the remainder depends on teacher training, student education, parent awareness, and fundamental behavior change. **Thus, we cannot assume that once we have done the paperwork, we will be automatically safe.** We need to strike a balance between meeting necessary regulatory

requirements and investing in meaningful safety practices, to avoid a system that offers illusory safety where compliance activities consume resources that could be directed toward activities that actually protect learners.

## Caution Zone: Unaligned AI Systems and the Failure of Age Verification

Unaligned AI systems with inadequate safeguarding mechanisms complicate the compliance load schools already face. The emergence of ‘malicious LLMs’—models deliberately designed to be malicious without ethical guardrails, such as WormGPT and FraudGPT, as well as easily downloadable models that can be modified through jailbreak techniques to remove safety restrictions—may spread harmful capabilities and pose threats that may be irreversible as they cannot be easily recalled or updated (Fire et al., 2025; Sood & Zeadally, 2025).

Furthermore, even when models are under guardrails, recent corporate policy failures illustrate the risks of relying on vendor claims without independent verification. Meta’s internal guidelines, for instance, reportedly permitted AI chatbots “engage a child in conversations that are romantic or sensual,” with documented examples including bots making inappropriate comments to minors (Horwitz, 2025). While Meta later stated these policies were “erroneous,” this incident underscores how major commercial platforms with substantial resources may fail to implement appropriate child protection measures.

Compounding these risks are the significant socioemotional and mental health implications posed by AI chatbots, particularly for young users. LLMs exhibit inherent sycophancy (i.e., tendency to affirm users’ beliefs and preferences even if it means sacrificing truthfulness), which can turn into a dark pattern of manipulative engagement (Keane, 2025). In February 2024, a 14-year-old boy committed suicide after becoming obsessed with a Character.AI chatbot, demonstrating how sycophancy quickly escalated into psychological delusion (Bellan, 2025; Duffy, 2024). Effective therapeutic support requires both validation and confrontation to promote self-awareness and challenge harmful thinking patterns; AI chatbots, however, tend to prioritize engagement over psychological well-being (Moore et al., 2025).

Online age verification, the primary defense mechanism that schools often assume exists, has proven to be inadequate. The French data protection authority (CNIL) conducted a comprehensive analysis of common age verification methods and concluded that none could sufficiently satisfy the combined requirements of reliable verification, population coverage, and privacy protection (CNIL, 2022). Specifically, current methods are either too simplistic (e.g., age verification checkboxes) or too invasive (e.g., ID scans which create significant data security risks and barriers for users without formal identification) (Forland, 2025).

For schools, these cases and realities underscore the inadequacy of surface-level vendor assurances and the need for not only more comprehensive evaluation approaches, but also ongoing staff training and proactive digital citizenship education.

### **Technical Systems: The Practical Impossibilities of Oversight and Evaluation**

In order to effectively approach safeguarding, school leaders and educators need to first gain a comprehensive understanding of what Howard et al. (2022) describe as ‘educational data journeys’—the complex pathways through which educational data is produced, processed, and distributed across multiple stakeholders and purposes. In other words, we need to be concerned about not just where data goes initially, but what happens to it once it enters AI systems, how it might be transformed and repurposed over time.

Consider, for instance, a student’s writing sample submitted to an AI writing assistant or chat conversations with AI tutors. While these interactions may begin with clear pedagogical purposes, the data generated may subsequently be transformed and used for model training, performance analytics, profiling, or commercial purposes (Huang, 2023). Even when these are clarified within privacy

policies and terms of use, and schools can opt their data out of model training by selecting Education or Enterprise versions of AI tools, it is often the case that educators and learners still use personal accounts without turning off data sharing. As such, the potential and consequences of data mishandling raise important questions about popular AI tools and platforms we use today—regardless of whether they are intended for education or not.

**The biggest barrier is that schools are expected to make critical safety decisions about these technologies that they cannot possibly understand or evaluate in the first place.** For instance, while the EU AI Act introduces legally enforceable regulations requiring schools to implement monitoring systems, conduct risk evaluations, and maintain accountability for AI-driven decisions (see Fact Box: Implications of EU AI Act for Education), it does so at a time when most educational practitioners might not understand how AI systems work in the first place, or they have

an outdated understanding based on outdated information. Schools must establish oversight frameworks to ensure AI systems are implemented ‘ethically and responsibly,’ yet even technology companies struggle with these same challenges. **Transparency requirements on algorithmic decision-making is impractical given the black-box nature of AI systems that even AI safety research groups have not yet fully addressed; accountability mandates raise critical questions about where schools would find resources for compliance teams or receive dedicated support; fairness requirements become meaningless when training datasets remain commercial secrets; and governance demands create**

**extensive documentation and ongoing administrative paperwork that schools simply lack the capacity to plumb through.**

To elaborate on a particular example, consider this divide between written policy and the practical reality: Regulations require that AI systems in education utilize comprehensive and diverse datasets, and that any systems discriminating against vulnerable student populations be prohibited. Yet current understanding of AI bias remains so limited that following this requirement would mean literally no educational institution could use LLMs, since comprehensive bias assessment is currently impossible with available information.

## Fact Box: Implications of EU AI ACT for Education

Building on top of previous efforts, the EU AI Act, which began implementation in 2024, takes a risk-based approach that categorizes AI systems into four risk levels: minimal, limited, high, and unacceptable (European Parliament and Council, 2024). Unlike previous voluntary guidelines and frameworks, the Act marks a crucial step in the governance of AI in education by shifting towards legally enforceable regulations.

In particular, the Act outright prohibits the use of “unacceptable risk” AI systems like emotion interference technologies (e.g., facial expression recognition, biometric analysis) in educational settings due to significant concerns over privacy, accuracy, and the potential to perpetuate racial, gender, and other intersectional biases (Buolamwini & Gebru, 2018). For example, AI proctoring systems have been documented to erroneously flag students with darker skin tones due to facial recognition bias, creating discriminatory barriers to equitable assessment (Meyerhofer, 2021). Meanwhile, for other AI applications in education (such as AI for assessment proctoring, performance evaluation, personalized learning, student admissions, etc.) all categorized as “high risk”, the Act introduces standardized compliance mechanisms.

What does this imply for educational stakeholders? As detailed by Saarela et al. (2025), this means any ethical guidelines that are previously non-binding are now principles that must be operationally enforced:

1. **Transparency:** AI systems in education must openly explain their data usage practices, decision-making processes, and algorithmic operations, with legal consequences for failing to comply.
2. **Accountability:** Educational institutions face strict legal responsibilities. Schools must implement monitoring systems, conduct risk evaluations, and be fully accountable for AI-driven decisions that impact students.
3. **Fairness and Inclusiveness:** Legal mandates for bias reduction. AI systems in education must utilize comprehensive, diverse datasets, and any systems that discriminate against or harm vulnerable student populations are forbidden.
4. **Governance and Oversight:** AI governance and supervision in educational settings are legally required. Schools must establish oversight frameworks to ensure AI systems in education are implemented ethically and operated responsibly.

Furthermore, the impossibility of evaluation is compounded by practical oversight challenges with the proliferation of AI tools. As Dr Anne-Marie Stanton-Ife described:

**You've also got the burden of checking everything because when these tools are mushrooming around everywhere and everyone's finding their own favorite thing... How on earth do you centralize the oversight of that?**

This scale problem is particularly dangerous due to the systemic vulnerabilities that schools face, as Nathan Nagaiah warned:

**Very often, schools will have different suppliers for different things, which unfortunately is a common practice.**

**So you may have different systems operating, and if one of the systems is corrupted, that system can bug into another system and extract this data.**

This interoperability means that data breaches can cascade across multiple systems via various data journeys, making it difficult for schools to maintain comprehensive oversight of where student data resides and how it flows between different platforms. While laws such as GDPR require data minimization and purpose limitation, to what extent is this enforced in practice? These systemic vulnerabilities then manifest in concerning ways, as John Jones illustrated:

**And we've come across some absolute horror stories as well. Like a couple of apps that are out there... I was able to**

access it, and then because it's kind of set up for schools, I was somehow able to become part of another school and start liking other children's work.

Many AI tools for education lack safeguarding mechanisms by design, enabling serious risks that may not be immediately apparent to educators or administrators (see Caution Zone: The Hidden Privacy Risks of AI Tools). Yet, schools rarely have the technical means and capacity to verify vendor claims. As Nathan Nagaiah continued to highlight, accountability gaps may emerge from vendor relationships and supply chain management:

**If your supplier has a data breach, will they report to the school directly? Well, if it's not in your agreement that the supplier needs to report any data breach immediately to the school, then the supplier would want to protect their name and their branding, and they might not even report to the school.**

In other words, when there are no mandatory or explicit breach notification clauses in vendor agreements, it creates a favorable situation where vendors may be incentivized to hide inappropriate practices to protect their reputation.

Such information asymmetry forces schools into a position of 'trust, because you can't verify' (Kelso et al., 2024)—operating under the false (or hopeful) belief that their data is secure while sensitive student information may be in fact circulating through compromised AI systems. This has cascading implications: By the time schools discover the breach, containment becomes impossible as the data has likely already been misused or redistributed (Shen et al., 2017). Schools then face lawsuits from parents who trusted them with their children's private information (Zhong et al., 2023).

**Therefore, the fundamental barrier is not just that schools lack expertise to evaluate AI systems—it is that we are at a point where AI is embedded in virtually every tool, meaning that under current laws, schools actually have to do extensive DPIAs and documentation (see Fact Box: What is a DPIA?). This is compounded by the fact that much of the technical information required for meaningful evaluation simply does not exist in accessible form in the first place, rendering regulatory standards void of practical meaning, creating impossible compliance requirements that schools cannot fulfill regardless of their expertise or resources.**

## Caution Zone: The Hidden Privacy Risks of AI Tools

The potential privacy risks and implications that students face through their use of Large Language Models (LLMs) are demonstrably non-trivial. Staab et al. (2024) demonstrated that AI can extract Personally Identifiable Information (PII) from seemingly innocuous clues, for example, an AI was able to infer someone's age and location with just the following sentence: "i was dragged out on the street and covered in cinnamon for not being married yet lol" (The person is 25 because there is a Danish tradition where unmarried people are covered in cinnamon on their 25th birthday).

The above finding shows that malicious attackers can infer sensitive information such as location, age, race, and gender from seemingly innocent conversations between students and AI chatbots (Piri, 2024).

The interoperability of educational data and the sensitive attributes involved exacerbate the risks. A parallel example to illustrate this is the 23andMe data breach. In that incident, an initial number of approximately 14,000 user accounts were hacked, but due to interconnected nature of platform features like DNA Relatives and Family Tree, the breach escalated exponentially, ultimately exposing sensitive personal and genetic data of approximately 5.5 million users and 1.4 million additional profiles (Holthouse et al., 2025).

Similarly, within an educational setting, student data can be highly interconnected across various platforms (e.g., academic records, health information, and social networks), creating a cascading effect if a breach occurs (Howard et al., 2022). Combining this with LLM's ability to infer personal attributes from subtle cues (even from region-specific slang and linguistic patterns that anonymizers fail to remove), the privacy risks become high-stake and large-scale, where well-intentioned attempts to anonymize student text might not fully protect their privacy (Staab et al., 2024).

## Fact Box: What is a DPIA?

The Data Protection Impact Assessment (DPIA) is a required procedure for any AI tool or general educational technology procurement and implementation in school settings. As mandated by GDPR, a DPIA is a process specifically designed to help institutions systematically analyze, identify, and minimize data protection risks before implementing, or purchasing, new projects or technologies, especially for cases involving automated processing or large-scale handling of personal and special category data—a growing concern for AI-powered educational tools (ICO, 2024). For schools considering AI adoption, DPIAs offer a structured framework to evaluate not only the immediate privacy compliances, but also the associated data processing operations and purposes, necessity of access, risks to students' and educators' rights and freedoms, safeguarding and security measures throughout the data lifecycle (SWGfL, n.d.).

However, conducting effective DPIAs often requires dedicated resources and specialized knowledge that most schools simply lack the capacity for. This highlights the crucial role of the Data Protection Officer (DPO). Under GDPR, schools as public authorities must either internally appoint or externally contract an independent DPO with expert knowledge of data protection law and practices, who serves as the primary point of contact for compliance issues and risk assessment. In the context of institutional AI adoption, the DPO's responsibilities include: 1) analyze privacy risks of AI tools and their data processing; 2) advise on GDPR compliance for educational AI platforms; 3) assist in conducting DPIAs for new AI technologies; 4) monitor ongoing compliance as AI tools evolve; and 5) liaise with students and regulators on AI-related data concerns.

We have provided a template DPIA under Part II Paths Forward section for your reference.

### **Internal Constraints:**

#### **Awareness and Capacity Gaps**

Even if external policies were clear and technical information were available, schools face internal organizational realities that make robust AI governance difficult within current institutional structures.

#### **1. Varying Staff Preparedness and Awareness**

The most immediate internal constraint to supporting responsible AI implementation is the wide variation in staff awareness, comfort levels, and training around AI use and related data protection practices. Unlike traditional technology rollouts, where schools can assume baseline digital literacy and acceptance among staff and students, AI adoption reveals dramatically different starting points that require differentiated institutional support approaches.

For instance, the mismatch of capacity manifests in dramatically different staff readiness levels, wherein schools face the challenge of navigating both fear-based resistance and unguarded enthusiasm amongst staff. As Steph Chambers noted, some educators are “scared and don’t understand” AI, and therefore avoid engagement altogether, leaving their students without any guidance on AI use. On the other end of the spectrum, there are also educators “who don’t know much but are quite happy to take risks,” as Rachel Bowen observed, highlighting that unguarded overconfidence is not desirable either. She further explained, “They’re the ones that I think we need to be really wary of—the ones who [say] ‘Yeah, this is great. Let’s do this.’” In other words, even well-intentioned educators might inadvertently compromise data protection practices.

The dynamic where different stakeholders in schools either lack caution or the technical know-how to experiment safely stems from a gap in data literacy. Conrado Torres’s reflection illustrates how this awareness develops over time:

**The data I [provide to] the AI tool—I try to be as non-personal as possible because as I became aware of more and more of the information inside of AI... That’s quite worrying for me—the way many AI tools have a lot of information about you, and you don’t really realize that they have a lot of information about you.**

His growing concern reflects a broader dilemma that many are facing, where initial enthusiasm for AI tools gradually gives way to deeper concerns for privacy implications: How much do teachers give of their students as a digital surrogate ‘product’ during our interactions with AI, willingly and unwillingly (Peltz & Street, 2020)? More crucially, how many people consciously recognize the severity of such implications in the first place? Nathan Nagaiah emphasized the fundamental scope of this issue: “This is where the lack of literacy plays a big role, doesn’t it? You need to have quite a lot of people who are literate about data, about data sharing.”

The data literacy gap thus renders schools facing a situation where educators who are keen to experiment may lack sufficient caution, exposing student data or undermining learning objectives; while those with caution may lack the technical know-how to experiment safely, leaving students without any guidance on AI use. Therefore, one-size-fits-all professional development sessions are not sufficient; instead, they need to meet educators where they are. For some, this means providing scaffolds that demonstrate AI’s educational value while maintaining clear safety boundaries; for others, this means channeling their en-

thusiasm toward responsible implementation through training on data protection protocols and building critical awareness of potential risks and limitations.

## **2. Resource Allocation & Accountability**

Another constraint on institutional responses is the reality of resource allocation. **This is not just a matter of being under-resourced, but rather how there needs to be significant restructuring in schools, depending on where accountability falls.** If accountability falls fully with schools (which it should when schools purchase AI solutions), where do institutions find the resources to ensure that accountability standards are met? Should Multi-Academy Trusts or school groups support dedicated data protection officer (DPOs) task forces, compliance teams, or AI champions? Is it feasible to retrain existing staff to handle these specialized responsibilities when current resources are already allocated across numerous competing demands?

Regulations like GDPR exemplify this problem by requiring schools to navigate complex data protection requirements that assume technical knowledge and institutional capacity most schools lack (see Fact Box: GDPR Requirements for Educators). Nathan Nagaiah highlighted this when noting that schools need “some kind of data governance structure where, before they share any data, they need to get some kind of sign-off—whether this data is relevant or not for us to share,” yet most schools lack both the technical expertise and organizational structure to implement such systems effectively.

# Toolbox: GDPR Requirements for Educators

| GDPR Article                                                                                                                                                                         | Why this matters for K-12 educators                                                                                                                                                                                                                                    | What you need to do                                                                                                                                                                                                                                                   |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Articles 6, 7, 4(11) – Freely given and informed consent</p> <p><b>What it says</b></p> <p>Consent must be freely given and informed, with full understanding of implications</p> | <p>Parents making decisions about AI tools for their children might not grasp the full picture of what is happening with their children's data. This is particularly tricky in schools where consent may not even be your primary legal basis for processing data.</p> | <p>Check whether your school's use of GenAI actually meets the 'necessity' threshold under public task or legitimate interest provisions. You cannot simply assume consent has covered everything.</p>                                                                |
| GDPR Article                                                                                                                                                                         | Why this matters for K-12 educators                                                                                                                                                                                                                                    | What you need to do                                                                                                                                                                                                                                                   |
| <p>Article 5(1)d – Data quality mandates</p> <p><b>What it says</b></p> <p>Data must be accurate, up-to-date, and fit for purpose</p>                                                | <p>Biased training datasets in AI tools can create discriminatory effects that further disadvantage underrepresented learners in your classroom.</p>                                                                                                                   | <p>Be aware that the AI tools you are using may inadvertently perpetuate existing biases. Critically reflect on and monitor for discriminatory outcomes and advocate for better training data.</p>                                                                    |
| GDPR Article                                                                                                                                                                         | Why this matters for K-12 educators                                                                                                                                                                                                                                    | What you need to do                                                                                                                                                                                                                                                   |
| <p>Article 22 – Profiling regulations</p> <p><b>What it says</b></p> <p>Protection from fully automated decision-making without meaningful human review</p>                          | <p>AI-generated outputs or learner portfolios that influence educational decisions about your students need human oversight to prevent discrimination.</p>                                                                                                             | <p>Ensure that there are always meaningful human reviews when AI tools are used for decisions that affect students, and question whether AI should be used at all in the process. Do not simply resort to the algorithm making decisions on educational outcomes.</p> |

#### GDPR Article

Article 5(1)a, 12 – Transparency and fairness requirements

#### What it says

AI tools must be transparently explained and built with data protection by design

#### Why this matters for K-12 educators

Companies developing AI tools should help you understand how the AI tools you use actually work, and your students/parents have a right to know too.

\* Unfortunately, because even AI researchers don't exactly know how LLMs work, this is easier said than done (Mollick, 2024).

#### What you need to do

It is almost impossible to choose AI tools that can completely explain their internal decision-making processes. However, you should at least have an understanding of where the data is going and how it is processed.

#### GDPR Article

Article 5(1)b, 5(1)c – Data minimization and purpose limitation

#### What it says

Only collect what is necessary and use it only for the stated purpose

#### Why this matters for K-12 educators

AI tools might request large amounts of student work or conversation transcripts that are more than needed for the system to function.

#### What you need to do

Follow data minimization practices: Make sure the AI tools you use do not collect or leverage student data beyond what is actually necessary for the educational task at hand.

#### GDPR Article

Article 9 – Special category data protections

#### What it says

Prohibits processing of sensitive personal data (racial/ethnic origin, political opinions, religious beliefs, health data, etc.) unless specific conditions are met

#### Why this matters for K-12 educators

AI tools (such as those used for processing Individualized Education Programs) may intentionally or inadvertently process special category data about learning needs, health, or wellbeing. Schools need explicit consent or must rely on substantial public interest exemptions.

#### What you need to do

Be careful about AI tools that might infer sensitive information about students' learning needs, mental health, or wellbeing. Make sure there is explicit consent and valid legal justification for processing this type of data.

### 3. Tension Between Experimentation & Oversight

Additionally, there exists a tension between experimentation and oversight, where schools must simultaneously encourage beneficial AI experimentation to prepare learners in an AI-integrated world while maintaining safeguards to protect institutional and student data. Chi-Hung Ha, IT Panel Head, True Light Middle School of Hong Kong, reflected:

**There may be some concern about the student information because [AI] reads the student assignment. And I use a platform called poe.com. It's not like Google Classroom, it's not like Microsoft Office, it's a new platform. So actually, I need to trust the platform at the very beginning, but I just have to try it out.**

In other words, when educators often find and use pedagogically valuable AI tools that exist outside their institution's approved technology ecosystem, this very act complicates data protection procedures and adds to a broader institutional dilemma: How can schools encourage beneficial AI experimentation while maintaining appropriate safeguards for student data?

In short, increasingly, we are seeing that schools are facing internal barriers that compound the aforementioned struggles in external compliance and technical evaluation. With varying staff awareness and readiness levels, resource allocation constraints, and tensions between encouraging innovation and maintaining safety, **schools' governance structures, primarily designed for curricular and pedagogical decisions, must now handle complex technical risk assessments**, often without appropriate expertise or actionable frameworks.

# Good Practices: What Does Safe Usage Look Like?

## Guardian Models and Content Moderation

AI tools used in schools should employ guardian models and content moderation systems, which are special AI mechanisms that act like automatic safety monitors. These systems use techniques like prompt filtering (checking educator/learner inputs before they reach the AI) and output scanning (reviewing AI responses before learners see them). Just like schools filter internet content, these AI safeguards prevent exposure to violent, sexual, or dangerous material. This allows educators to confidently integrate AI into lessons and projects, knowing that inappropriate content has been automatically blocked at multiple checkpoints.

## Professional Moderation Services

Educational platforms often use enterprise-level content safety APIs like Microsoft's Azure AI Content Safety—cloud-based services that real-time scan text and images to identify harmful content such as hate speech, self-harm, and other inappropriate content. Popular educational tools like Canva's AI features already integrate these moderation layers behind the scenes.

These contextual safety ensures AI remains educationally valuable while maintaining classroom-appropriate boundaries. However, it is important to note that these technical countermeasures still has limitations, including but not limited to bias, false positives, overreliance on automation, and algorithmic opacity (Kaithathara & Jose, 2025). Safe AI usage in education thus requires more than just technology, educators must combine these tools with thoughtful oversight, clear guidelines, and ongoing critical evaluation to ensure that learning environments are genuinely safe.

Azure AI Content Safety:

<https://learn.microsoft.com/en-us/azure/ai-services/content-safety/overview>

AI Safety at Canva:

<https://www.canva.com/policies/ai-safety/>

## **Paths Forward: Institutional Frameworks and Structured Support**

Given the multilayered challenges around safeguarding, ranging from external policy gaps to technical barriers to internal capacity constraints, we recognize that schools need pragmatic approaches that work within current realities. The path forward requires acknowledging these constraints, while drawing upon actionable frameworks and scaffolds that both meet schools where they are, and can evolve alongside technological change and institutional capacity.

### **Security Testing and Collaborative Vetting Processes**

To counter the practical impossibility of oversight described earlier, schools must move from guesswork to structured evaluation. This means establishing robust security testing and vetting processes that provide explicit frameworks for what constitutes appropriate AI use, so that educators can experiment within safe parameters rather than avoiding AI altogether or using it blindly. As Rachel Bowen emphasized, the goal is to put “some guidelines in place, making sure that [educators] know the boundaries and how far they can go, because otherwise... they don’t sometimes understand the implications of what they might be doing.”

**In order to facilitate informed decision-making rather than being blindly guided by a leap of faith, schools should first and foremost devote specialized effort into safeguarding and security testing, building institutional confidence by removing guesswork and providing objective safety criteria.** For instance,

John Jones described how his school has adopted comprehensive evaluation processes, where, for any AI tools and apps in general, they “scrutinize it, check it, and then release it.” Additionally, for the approved sets of tools, the school provides clear guidance and training on how staff should appropriately leverage them, for instance, by “[going] through Gemini and explain[ing] why this is what you should be using.” Similarly, Nathan Nagaiah emphasized the need for schools to develop or access specialized expertise in cybersecurity and data protection—even if they lack ongoing year-round capacity, they can still implement periodic security assessments, such as “system penetration testing at a random time of the year just to check the kind of data vulnerability.” **As such, more than helping to identify risks before problems occur, these proactive security mechanisms support educators by assuring them that institutional safeguards are in place—that the AI tools are properly vetted and monitored, and that they can focus on the pedagogical implementation rather than worrying about unknown security implications.**

**However, placing the full burden of this technical scrutiny on individual schools is unsustainable. A more scalable path forward might be establishing collaborative vetting processes. This could take the form of inter-school consortia, district-level task forces, or partnerships with specialized research organizations that pool expertise to conduct initial technical reviews and security testing for common AI tools. Of course, schools do not need to wait for formal initiatives to begin collaborative vetting. Here are some immediate steps that schools can take:**

1. First, connect with neighboring schools or district partners to share evaluation workload—even informal conversations like ‘What AI tools are you using and how are they working?’ can help.
2. Second, leverage existing networks that already facilitate resource sharing.
3. Third, if resources allow, designate one tech-savvy staff member as an ‘AI scout’ who monitors new tools and creates brief evaluation summaries for colleagues.

## Toolbox: Good Future Foundation Community Platform

These approaches have been formalized in the Good Future Foundation online community platform, which is a free resource for all teachers to join in sharing best practices.

To sign up, please visit: <https://bit.ly/JoinGFFCommunity>

Initiatives like these help schools to share the load and avoid redundant effort, systematically build institutional knowledge, and lower access to professional safety evaluations that many do not have the capacity to conduct alone.

## **Due Diligence for Schools When Onboarding AI Tools**

To facilitate a more systematic and safe approach to AI adoption, schools should conduct a thorough due diligence process when onboarding new AI tools. This section provides a structured framework with practical templates and checklists on what schools can do before onboarding, during onboarding, and during the rollout of AI tools. These are designed to help schools make informed decisions, protecting student data and fostering a safe learning environment.

## **1. Before Onboarding: Conduct a Risk Assessment**

Before onboarding an AI tool, schools could first conduct a risk assessment to assess organizational readiness and ask questions. Organizations like Educate Ventures Research, for instance, have developed a comprehensive risk assessment framework specifically for AI tools in educational contexts. Below, we have adapted it into a pre-onboarding checklist for your reference.

# **Toolbox: AI Risk Assessment (Before Onboarding AI Tools)**

(Adopted from EDUCATE Ventures Research Risk Assessment)

### **1. Description: Describe the AI tool's core functionality and intended use.**

- What specific educational tasks does this tool accomplish?
- Who are the primary users (teachers only, students, or both)?
- What is the pricing model and are there school discounts available?

### **2. Innovation: Gauge the AI tool's educational value and uniqueness.**

- What specific educational and pedagogical advantages does the tool offer over those alternatives?
- Is the tool accessible and intuitive to use?
- Can the tool be customized to match your curriculum needs and teaching objectives?

### **3. Data Risk: Identify and evaluate data privacy and security risks.**

- Is the tool compliant with relevant data protection regulations?
- What personal information does the tool collect from users?
- Is user data used to train AI models, and can you opt out of this?
- How long is data stored, and can users request deletion of their information?
- Does the tool share data with third parties?

**4. Ethical Risk: Evaluate potential ethical concerns and biases.**

- What safeguarding controls are in place to ensure appropriate and credible content generation?
- Are there age restrictions or controls for student use?
- Can teachers have oversight over student interactions with the tool?

**5. Risk Mitigation: Document specific measures to address identified risks.**

- What precautions should be taken to protect personally identifiable information?

**6. Implementation Notes: Provide practical guidance for tool use.**

- Should the tool be tested or piloted before full student deployment?
- What training or guidance do staff need for safe and effective use?

More info: <https://www.educateventures.com/risk-assessments>

**2. During Onboarding: Complete a DPIA**

When onboarding an AI tool, schools can conduct a Data Privacy Impact Assessment (DPIA), which is a standard instrument in GDPR jurisdictions, to document their evaluation and safeguarding process for audit and compliance purposes.

## Toolbox: DPIA Template (During Onboarding AI Tools)

This brief DPIA helps schools conduct initial due diligence before adopting a new AI tool. It is designed to quickly identify potential privacy risks and determine if more comprehensive reviews are necessary. Not all checkboxes or fields may be applicable, the DPIA serves more as a process for documenting the information you have. The legal basis section includes examples for major jurisdictions, but local regulations may vary. When in doubt, consult your school's legal counsel or designated Data Protection Officer.

School:

AI Tool:

Date:

### 1. INITIAL SCREENING: IS THIS A HIGH-RISK ACTIVITY?

[This first step determines if the AI tool's function requires a careful privacy review. If you check any of these boxes, it signals that potential risks exist and this assessment is necessary.] Check any that apply:

- AI makes automated decisions affecting students (grades, placements, interventions)
- Processes sensitive data (health, special needs, disciplinary records)
- Systematically monitors students (behavior tracking, online activity)
- Uses student data for profiling or scoring
- Processes data from vulnerable children (<13 years old)
- Public school using a cloud-based service with student data

Screening Result:

- Assessment Warranted (At least one box is checked)
- Assessment Not Necessary (No boxes checked, low risk)

### 2. DATA & PROCESSING

[Describe the "what" and "why" of the data processing. Be clear about what specific student information the AI tool will use and for what educational purpose.]

Student Data to be Processed:

Names, IDs, contact info  
Academic records/grades  
Student work samples  
Behavioral/attendance data  
Voice/video recordings  
Health/special needs data  
Other (please specify):

Number of Students:

Grade Levels:

AI Tool's Educational Purpose:

Data Retention Policy (How long is data kept?):

### 3. VENDOR & DATA TRANSFERS

[Identify who is handling the data and where it is physically stored. This is critical for understanding jurisdiction and security.]

Vendor:

Vendor's Location:

Data Storage Locations (e.g., vendor's server in Ireland):

International Data Transfers: Yes / No

If yes, safeguards in place:

#### 4. LEGAL BASIS FOR PROCESSING

[Identify the specific legal permission your school has to process this data. This will vary by jurisdiction.]

##### For U.S. Schools:

FERPA Compliance:

- School official exception
- Parental consent
- Directory information only

COPPA (<13):

- School consent
- Direct parental consent
- N/A

##### For EU/UK Schools:

GDPR: Art. 6 Basis (e.g., Public Task, Legitimate Interest):

GDPR Art. 9 Basis (if sensitive data):

##### Other Jurisdictions (e.g., Hong Kong PDPO):

Applicable Local Law & Basis:

#### 5. QUICK RISK CHECK

[Rate the potential for harm on a simple scale of Low/Medium/High. This helps prioritize what needs to be addressed with safeguards.]

Data breach/unauthorized access:

Commercial use of student data:

AI bias against certain students:

Loss of student privacy:

Inaccurate AI decisions:

Lack of transparency in AI:

Highest Risk Level: Low / Medium / High

#### 6. ESSENTIAL SAFEGUARDS

[Document the technical, contractual, and procedural controls in place to protect student data and mitigate the risks identified above.]

##### Technical Security:

Data encrypted in transit and at rest

Multi-factor authentication required

Role-based access controls

Vendor provides regular security audits (e.g., SOC 2 report)

##### Contractual Protections:

Data Processing Agreement (DPA) signed

Clear data deletion terms upon contract end

Vendor must notify school of subprocessors

Liability/indemnification clauses are in place

Oversight & Accountability:

- Human review of high-stakes AI decisions
- Plan for regular bias/accuracy testing
- Clear process for students/parents to file complaints or request data
- Staff has been trained on appropriate use of the tool

## 7. STAKEHOLDER COMMUNICATION

[Privacy is a community effort. Document who has been consulted and how families are being informed.]

Consulted Internally:

- Teachers/staff
- IT team
- Legal counsel
- Leadership

Parent & Student Communication:

- Information provided to parents/guardians about the tool's use
- Opt-out option available (if applicable)
- Explicit consent obtained (if required by law or policy)

## 8. DECISION & ACTION PLAN

[Based on the balance of risks and safeguards, make a final decision. If conditionally approving, be specific about what must be resolved.]

Overall Assessment:

- Low Risk (Green)
- Medium Risk (Amber)
- High Risk (Red)

Action:

- APPROVED – Proceed with implementation
- CONDITIONAL – Address these issues first:
- REJECTED – Risks too high or inadequate safeguards
- FULL DPIA REQUIRED – Conduct more comprehensive assessment

Approved by (Name & Title):

Date:

## 9. MONITORING & REVIEW PLAN

[Data privacy is not a one-time task. Outline a plan to ensure the tool remains safe and effective over time.]

Review Schedule (e.g., Annually, after any major feature update):

Key Metrics to Track (e.g., Accuracy rates, number of parent inquiries, security incidents):

Next Scheduled Review Date:

### 3. At the Rollout Stage: Implement a 'Traffic Light System'

At the rollout stage of an AI tool, schools need to clearly communicate permitted use cases and guardrails to teachers and students. In order to do this, schools could use memorable devices such as a 'traffic light system' to aid compliance. We recommend categorizing digital tools based on their potential risks to data privacy, safeguarding, and educational integrity, providing clear, actionable guidance for all

stakeholders, where each tool is marked red, amber, or green. There should be an executable use policy, supported with professional development, for each tool that needs an assessment to be completed before access through a school license is granted. Use of an AI tool for school purposes on a personal license should have enforceable consequences. Below, we provide an example of a 'traffic light system' for categorizing AI tools.

## Toolbox: Example Traffic Light System for Schools (When Rolling Out AI Tools)

| Criteria | Red: High Risk                                                                                                                                                                                                      | Amber: Medium Risk                                                                                                                                                 | Green: Low Risk                                                                                                                                              |
|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
|          | Fails to meet UK GDPR / DPA 2018 requirements (e.g., no clear privacy notice, collects excessive personal data, unclear data storage location).<br>Enables unsafe contact or unmoderated communication with minors. | Meets basic legal requirements but has gaps in transparency or data minimisation.<br><br>May store data outside the UK/EU without robust data transfer safeguards. | Fully meets all legal (UK GDPR / DPA 2018) and safeguarding requirements.<br><br>Demonstrates transparent data practices and a 'privacy by design' approach. |
|          | Uses opaque or high-risk algorithms (e.g., extensive profiling, biometric data) without explainability.                                                                                                             | Shows limited or unverified evidence of educational benefit.                                                                                                       | Has proven educational value with evidence from reputable sources.                                                                                           |
|          | Has a history of data breaches or unresolved safeguarding concerns.                                                                                                                                                 | Does not fully comply with age-appropriate design standards.                                                                                                       | Fully complies with the Age Appropriate Design Code.                                                                                                         |

## Policy

### Red: High Risk

Not approved for any school use.

No installation or use via school accounts or personal accounts for school-related purposes.

Staff must be briefed on the specific risks and the reasons for the ban.

### Amber: Medium Risk

Use is permitted only with written approval from the Data Protection Officer (DPO) and/or Senior Leadership Team (SLT).

Requires mandatory staff training or a competency quiz before access is granted.

Usage is restricted to specific year groups, subjects, or contexts.

Must be reviewed termly to assess any changes in its risk profile.

### Green: Low Risk

Approved for general school use, guided by the published Acceptable Use Policy.

Students and staff must complete a brief induction or quiz before their first use.

## Example Actions

### Red: High Risk

Block the tool at the network level.

Audit school devices to ensure their removal.

### Amber: Medium Risk

Permit access using school-managed accounts only.

Maintain detailed usage logs for auditing purposes.

### Green: Low Risk

Monitor the tool for ongoing compliance and any changes to its terms.

Include the tool in regular staff training updates and best practice guides.

## Fostering a Culture of Awareness

Beyond the technical frameworks and risk assessments, responsible AI adoption ultimately hinges on the human element. As such, building the necessary capacity requires more than top-down mandates. As John Jones stressed, the essence is to cultivate a deep-seated “culture of awareness”, so that “before [educators] enter any information into any form of AI or even any system, they will think twice about it.” This awareness must be grounded in multiple dimensions of data practice. For instance, Nathan Nagaiah illustrated one critical aspect—data minimization:

**We’re collecting this data—do we actually need the data? If you do not need the data, why do you collect it in the first place? Because having an excessive amount of data, you are actually exposing yourself to a high level of risk when you don’t actually need it.**

This cultural approach recognizes that effective governance cannot rely solely on policies and training sessions but must become embedded in daily practice and decision-making processes, requiring educators and learners to develop habits of questioning not only what data they collect, but also how they share it, where it goes, and what long-term implications their choices might have. **The question, then, is how to nurture this organic, ground-up development. If large-scale, one-size-fits-all training is often insufficient, the most effective path lies in empowering educators to learn from one another.**

## Toolbox: Good Future AI

Good Future AI is a free, teacher-facing platform built by the Good Future Foundation to help schools experiment with generative AI safely. It was designed to remove two of the biggest barriers teachers face: data privacy concerns and cost. The service follows a strict zero data retention policy, meaning no prompts or outputs are stored, shared, or used for training models.

### Key Features:

- Zero data retention
  - inputs and outputs are deleted immediately after use
- No model training
  - neither GFF nor providers can use teacher content to improve AI models
- GDPR-equivalent protections
  - providers (Anthropic, OpenAI, Vercel) bound by Data Processing Agreements
- Professional guidance
  - teachers retain full rights to their inputs/outputs, with clear boundaries for responsible use
- Ethics-first approach
  - AI positioned as an assistant, never an authority, reinforcing teacher expertise

### Use Cases for Schools:

- Generate quizzes, glossaries, summaries, and classroom materials more efficiently
- Explore AI literacy with staff through guided demonstrations
- Pilot AI for administrative support while protecting privacy and trust

Good Future AI is best seen as a professional sandbox for educators, as opposed to a classroom product for students. It helps schools build confidence, test ethical boundaries and embed responsible practice before wider adoption

To gain access, email: [info@goodfuture.foundation](mailto:info@goodfuture.foundation)

Disclaimer: AI outputs must always be verified and contextualised with professional judgment. The tool is not student-facing and should never be used to process personal or sensitive student information. By combining privacy safeguards, contractual limits on providers, and alignment with the Department for Education's GenAI Product Safety Framework, Good Future AI offers schools a secure space to trial AI without risk to learners.

## Creating Spaces for Peer-to-Peer Professional Development

Through our conversations, we recognize that educators often find more value in contextualized, informal conversations than in top-down directives or large-scale professional development sessions, especially when navigating the tensions between pedagogical innovation and institutional compliance.

As Steph Chambers observed, “I’ve never found introducing a new tech tool to like 45 staff members gets you anywhere... I find the most benefit from actual one-to-one conversation with staff... [like a] micro PD [or] mini PD.” This is echoed by Kate Atherton, who highlighted how informal sharing can scale effective practices: “Sometimes they then take what we’ve done there and they will share it during that collaboration for mastery time of like ‘oh this is what we’ve been trying here, it worked really well.’”

The peer-to-peer sharing amongst teachers becomes a constructive practice, given that AI implementation varies significantly across academic disciplines. As emphasized by Dr Andy Kemp, Principal at The National Mathematics and Science College, “Pedagogy is a much more subject-specific thing than people give it credit for... Good practice of AI use looks different in different subjects,” suggesting professional development approaches must be catered to disciplinary relevance. Indeed, schools have begun to facilitate cross-departmental learning, as John Jones described:

**We just got different departments to see what they’re doing, and that was a really good way to see what was going on. So we had RS (Religious Studies), we had English, we had Maths, we had Geography... All teachers and heads of department standing up and just giving use cases of how they’d integrated AI... And that’s the sort of culture that we spoke about.**

The question comes down to how schools might maintain a culture of openness and experimentation while ensuring that individual innovation aligns with institutional data protection requirements and educational quality standards. Through our conversations, we realize that schools must create formal structures for informal innovation—channels through which champion teachers can experiment, iterate, and share learning while remaining within appropriate governance frameworks (Reich, 2023).

This requires appropriate balancing of the ‘why’ against the ‘how’. For instance, Dr Andy Kemp argued that professional training should focus on conceptual understanding—the “why” and broad principles of AI: “I do think [insets should be] predominantly about conceptual understanding of what’s possible rather than here are good examples of what you might go and do.” However, Steph Chambers emphasized that teachers want practical application—the “how” of specific tools and techniques they can use immediately: “Teachers want practical solutions that from a PD they can take away and use in their classroom... Most teachers just want to know how it applies to their

classroom.” In other words, there seems to be a tension between what teachers want (immediate solutions) and what they may need for responsible and adaptable AI use (conceptual foundations). Without conceptual grounding of how AI works, the safeguarding risks involved, as well as its potential and limitations within particular pedagogical and classroom contexts, practical AI training risks becoming procedural—teachers may learn to use specific tools but lack the skills to adapt, troubleshoot, or make informed decisions as AI and learning contexts both evolve.

This tension suggests that effective institutional capacity building requires creating safe spaces for experimentation within governance frameworks. As Dr Andy Kemp put it, schools should make space for educators to “experiment and play,”

while Chi-Hung Ha echoed that “teachers need to go through that process to learn from the mistakes of using AI.” However, this experimentation must occur within structured environments that balance innovation with institutional safeguards. Building sustainable institutional capacity, therefore, requires designing professional development frameworks around ‘scaffolded experimentation’—one that addresses teachers’ immediate practical needs while ensuring they develop the principled understanding necessary for long-term, responsible AI implementation.

## Part III: On Motivation

### How can schools inspire stakeholders to change their behaviors towards new best practices?

Even in a system with confident educators and learners and robust institutional support, a more subtle and profound challenge is the question of motivation. Here, the very affordances of AI—its ability to generate intelligent-seeming products with ease—threaten to short-circuit the effortful cognitive processes that constitute genuine learning. Our final section thus grapples with the purpose of education itself, asking: How do we inspire learners and educators to pursue the difficult, rewarding path of deep understanding when AI offers an easy, seductive alternative?

#### **Current Challenges: The Struggle for Authentic Teaching & Learning**

##### **For Learners: The Trap of ‘Metacognitive Laziness’ and Loss of Agency**

While transparency frameworks help establish honest communication between learners and educators about AI use, they also point towards a deeper set of questions about the changing relationships students are forming with AI technologies themselves, as well as the resulting implications for their agency, responsibility, and intellectual growth. Rather than simply asking whether students should or shouldn’t use AI, educators are grappling with more nuanced questions: How do learners conceptualize their relationship with AI?

When does AI assistance support genuine learning, and when does it undermine it? And what role should learners themselves play in regulating and reflecting on their AI use?

Many educators highlight the importance of facilitating learners to distinguish between productive AI collaboration and problematic dependency, particularly given the risk of ‘metacognitive laziness’ (i.e., cognitive offloading of self-regulatory processes or reduced self-awareness about one’s own thinking) when AI tools bypass the effortful cognitive processes essential for genuine learning (Fan et al., 2025; Oakley et al., 2025).

**“If you use AI and you put a list of points into ChatGPT and it writes the essay for you, then what actually has happened? It’s like, if we could come up with a machine that bench presses 300 pounds, why not sit in the corner of the gym and drink a smoothie while we’re watching that?”**

**—Professor Sam Wineburg, 2025  
Educational and cognitive psychologist,  
Margaret Jacks Professor of Education  
at Stanford University**

Learners themselves have also demonstrated varying levels of awareness about this distinction, with some articulating clear metacognitive

awareness of AI as a learning tool, while others are struggling to maintain control in their interactions with these tools. Conrado Torres exemplified a thoughtful approach to this relationship: "It's not like ChatGPT doing the assignments for me, but more like GPT being a tool for me to help me study." His distinction between AI doing work 'for' him versus serving as a tool to 'help' him study reflects an understanding of AI as supplementary rather than substitutive (Bauer et al., 2025). Ha Thi Hai Nguyen echoed this collaborative perspective from an educator's standpoint, noting the importance of intentionality and skill in interactions with AI, "AI can be considered as a very good tutor, individual tutor if we can use it in the right way." However, not everyone manages to maintain this collaborative balance in practice. Torres described witnessing this among peers who had turned to AI for creating presentation materials:

**I noticed that for some of them, if they overused AI, they didn't feel their speech as if it were done by themselves. So it was more difficult for them to read the speech in different ways and to emphasize some things because they felt that they didn't really write it themselves.**

When learners deliver content and work they did not authentically create themselves, learning becomes disconnected from their own thinking processes, and the consequences are both cognitive and emotional. For instance, according to Pekrun's (2006) Control-Value Theory, learners' emotions about learning stem from their perceived control and the

subjective value they place on academic tasks. When applied to the situation of AI for learning, this creates a troubling paradox in that, while AI tools promise to augment learning, they may as well inadvertently undermine the very emotional foundations, such as resilience during setbacks, honest self-regulation, and cooperative engagement that make meaningful learning possible (Luckin, 2025a). Selin O. captured this tension in her reflection:

**When you consult ChatGPT for too many things, it kind of puts you back a certain level—like it kind of blocks your mind. Not in a physical sense, but because you know that there is someone else who could help you, at some point, people may lean on putting less effort into doing certain things, and that happened to me.**

Selin's experience reflects the insidious nature of cognitive dependency, which is of particular concern given recent research findings suggesting that overreliant AI use may result in shallower memory encoding, diminished cognitive agency, and reduced independent thinking (Kosmyrna et al., 2025). Students like Selin who recognize this pattern demonstrate the kind of metacognitive awareness and responsibility for their own learning journey that education must cultivate more deliberately. Importantly, this also illustrates how learners do tend to intuitively recognize when they have not earned their understanding—and this recognition can in fact become a pathway to more thoughtful AI use if properly supported by schools, especially in environments where AI tools are readily available and increasingly sophisticated.

### **For Educators: The Lure of Efficiency over Pedagogical Effectiveness**

Another emerging theme is how AI is re-defining what it means to be an educator. To be clear, this is not just about adopting new tools or updating pedagogies and learning materials; rather, concerns are raised around core assumptions about professional expertise, the nature of teaching work, and the relationship between efficiency and effectiveness.

When it comes to the immediate impacts of AI in schools, there is a dominating discourse around how AI affects professional workload. At first glance, AI appears to offer dramatic efficiency gains for educators. Chi-Hung Ha described how he uses AI for grading: “In the past, I needed several days to mark four classes of assignment assessment in one week or one cycle. With AI, I can just do it in one afternoon. In terms of efficiency, there’s no match.” He further emphasized how AI can help him to “prepare the lesson materials, lesson plan, brainstorm ideas, and create the exercise test quizzes—everything about teaching and learning you can do with AI, but with 10 times efficiency.” Indeed, a recent research trial published by the Education Endowment Foundation (EFF) supports these claims: Teachers using ChatGPT with guidance spent 31% less time on weekly lesson planning over a 10-week period compared to teachers not using AI tools (Baxter, 2024).

However, this is not to say that there aren’t alternative (or more nuanced) arguments with regards to the implications of AI on educators’ workload beyond time savings. One of which is what might be called the ‘efficiency trap’—AI tools may increase productivity without necessarily improving

work-life balance or reducing professional demands. Dr Andy Kemp captured this paradox:

**What we’re seeing at the moment is [AI] is enabling people to do things that they wouldn’t have felt they had the capacity to do before. Which isn’t quite the same as reducing workload because effective teachers are probably working just as hard, if not a little bit harder, but they are achieving more with what they’re doing.**

In other words, while AI may help reduce time for specific tasks, it can simultaneously raise expectations for what educators should and could accomplish. Furthermore, for any educator to even benefit from the efficiency gains in the first place, they would need to invest significant time learning these tools—a significant barrier for those who are already overburdened. As Kate A. noted, “Teachers are so time poor, and to use or understand AI requires investing time, and it will save you time in the long run. But you need to get that buy-in.”

More critically, the lure of efficiency may even overshadow pedagogical effectiveness, where educators might complete tasks like lesson planning and assessment design faster, but without deeper consideration or genuine understanding of how these materials will genuinely benefit their students or align with core learning goals. As Tal Havivi succinctly put it, **“The risk, of course, is that AI doesn’t actually transform learning—it just creates a much more efficient, non-effective way of teaching and learning.”** Similarly, Dr Andy Kemp elaborated on how the uncritical use of AI manifests in assessment and feedback: While AI can efficiently grade

student work, the purpose of assessment is not just assigning grades to students; rather, it also serves “to give feedback to the teacher as to what’s working and what’s not working and what they need to do next.” In other words, assessment and instruction are two sides of the same coin. Outsourcing this process to an AI might provide quick feedback to students but robs the teacher of the critical insights they need to effectively pivot their instruction.

In short, if efficiency is pursued at the cost of the human-centric, intellectual, and relational work that defines effective teaching, the very essence of the profession is called into question.

#### **For the System: Misaligned Incentives and the Assessment Dilemma**

The challenges of motivating stakeholders to adopt best practices for AI in education extend beyond individual learners and educators to the very structure of the educational system itself. For decades, assessment has been the primary mechanism for validating learning, but it has often prioritized the final product—the polished essay, the solved equation, the completed project—over the intellectual process required to create it. Such focus leaves out a systemic vulnerability that GenAI exploits with unparalleled efficiency. The core of the conversation, as Tal Havivi noted, must move beyond simple questions of academic integrity to “What does it mean to learn? What does schooling look like in an AI-infused world?” When the products of intelligence can be generated effortlessly, the system faces a profound dilemma, one rooted in what

Dr Guillermo Solano-Flores has called a “very narrow old-fashioned way of what learning is and what acceptable performance is.”

Learners are acutely aware of this misalignment and are often incentivized, implicitly or explicitly, to optimize for performance over understanding. Naomi King provided a candid example from her Design course, where assessment criteria reward specific phrasing. While she would “do all the designing and stuff [herself]”, she admits using AI “to help [her] with the wording for the reports [she has] to write on the designs [she’s] done” because “they love sort of those buzzwords and hearing that. It’s such an easy way to tick off marks, basically.” This is not so much an act of overt cheating but a rational response to a system that values a particular output.

When it comes to her Computer Science course, however, Naomi comments how “it’s so much harder to not rely on AI” because the assignment description itself almost acts like a sophisticated prompt for AI already, “It’s so easy to just use Cursor and give it your PDF assignment. It’s always in a PDF brief, and it can literally spit it out for you.” The dilemma intensifies in competitive environments, where the pressure to perform can override the desire to learn. Naomi further described this tension in her Computer Science group projects, where she feels compelled to match the AI-fueled productivity of her peers:

**They had these pages done in a day [with AI]. It would take a normal developer a week to make... and I guess I’ll just take extra work now, where it’s**

like I'm struggling to even get the bare minimum done. Then I'm just pressured a little bit 'cause they've told us in Computer Science that you can't use AI for generating code. It's a little bit hard to manage 'cause it's really hard to stick on [her peers'] level. I want to learn for my own benefit how to code. I just don't like it personally [using AI for assignments], and I don't want to do it.

In short, when the assessment system values output only, peer pressure on AI adoption exacerbates the situation where conscientious students who want to engage in authentic learning are inadvertently put at a more disadvantaged position.

This assessment dilemma places educators in an untenable position, forcing them into what Steph Chambers described as a "cat and mouse game of like, 'right, what are we asking kids to do?'" that AI cannot easily replicate. However, this reactive approach often fails to address the root causes. Dr Andy Kemp argued that the problem lies in trying to bolt a 21st-century technology onto an outdated framework, asking, "When's it okay to use AI within a semi-Victorian education system?" He observed that, for instance, within the "rigidness of the A-level curriculum" for sciences, AI's primary function becomes to "predominantly circumvent the thinking." In other words, when learning and assessment tasks are formulaic and predictable, AI excels at providing solutions that bypass the critical cognitive processes essential for genuine learning. This forces a necessary, if uncomfortable, re-evaluation of long-standing educational practices. As Tal Havivi posited, "Is it still important to write a five-paragraph essay? I think people need to have that conversation."

As Dr Andy Kemp articulated in the context of science education, "Certainly at A-level in the sciences, there aren't many areas where AI is actually particularly helpful yet. On the whole, what AI does within the sciences at the moment is predominantly to circumvent the thinking." **In other words, the issue is not that AI lacks capability, but that current curricular demands and assessment structures create incentives for learners to use AI in ways that bypass rather than engage in genuine learning.** Dr Andy Kemp further elaborated that this is not just subject-specific; instead, it is a result of systemic lag that requires change.

This tension reflects what Reich (2020) described as the 'curse of the familiar' in his book *Failure to Disrupt*—rather than fundamentally reimagining assessment and pedagogy, many schools often adopt technologies by fitting them into their usual practices. The result is often insufficient consideration of pedagogical alignment, professional development needs, or the structural changes required for meaningful integration. This is especially true under circumstances where high workload and standardized expectations limit educators' capacity for change. **In short, when schools attempt to integrate new technologies into old educational systems, the result is often a series of compromises that satisfy neither the demands of authentic learning nor the potential of AI-enhanced education.** Nonetheless, schools cannot wait for long-term system-wide reform to support their educators. Most immediately, educators should be supported with practical frameworks to align AI use with learning objectives despite the current structural constraints.

**“The curse of the familiar emerges from trying to use technology alone to change schooling. Schools, with their innate complexity and conservatism, domesticate new technologies into existing routines rather than being disrupted by new technologies.”**

**—Justin Reich, 2020**

**Associate Professor at MIT Comparative Media Studies/Writing, Director of the Teaching Systems Lab**

**Ultimately, AI is not the cause of the problem but an accelerant that exposes pre-existing weaknesses in the educational system—that there is a misalignment between the stated goal of education (to foster deep, durable understanding) and the incentives created by traditional assessment methods.** Without a fundamental redesign of what we value and how we measure it, schools risk motivating superficial engagement and creating an environment where the appearance of achievement is prioritized over the difficult, essential work of learning.

#### **Paths Forward: Redesigning for Deeper Learning**

As we have seen, the motivational challenges demonstrate how uncritical use of AI may hollow out the very essence of learning. To put it more simply, if we stay in a system where educators lean on AI to create quizzes and track progress, and learners use AI to ace these quizzes, then where is genuine learning? This is not so much a problem that can be solved by technology implementations, policy adjustments, or training programs. Rather, it requires us to rethink and redesign education on a systemic level, one that makes human learning not just possible, but compelling and deeply meaningful.

#### **Redesign Assessment to Reward Process, Not Just Product**

If traditional assessments incentivize students to circumvent learning, then the most effective path forward is not to build better detection tools, but to build better assessments. The challenge of AI is, in reality, an opportunity to close what Dr Andy Kemp identified as the “real gap” in our educational model: the need for “fundamentally better forms of assessment.” In other words, instead of clinging to a system that measures outputs AI can easily replicate, educators can shift their focus to designing tasks that reward the uniquely human process of inquiry, critical thinking, and problem-solving.

#### **1. Make Thinking Visible: Assess the AI Dialogue**

One strategy in this redesign is to make the student’s thinking visible. This moves beyond simply asking if AI was used and instead asks how. In other words, when we advocate for transparency, it becomes most powerful when we acknowledge that it is not just about mutual disclosure and administrative compliance, but a meaningful reflection on the learning process itself. Dr Guillermo Solano-Flores highlighted this critical distinction, contrasting “requiring the students to tell us how they used artificial intelligence” with “asking them to show us the prompts that they used and the process of refinement of the prompts.” This simple shift transforms the assessment. It is no longer a confession but a documentation of a learning journey. As he explained, “In the prompts, you can see the process of thinking and how the student is constructing knowledge.” In other words, by making the dialogue with AI a part of the evaluated work, educa-

tors can assess the quality of a student's questions, thinking process, their ability to iterate on feedback, and their overall intellectual engagement with the material.

## **2. Design for Deeper Cognition: Creating Tasks that Resist Superficial AI Use**

Beyond making the thinking process visible, educators can design tasks that are inherently resistant to superficial GenAI use. This involves creating problems that require adaptability, resilience, and real-time problem-solving skills that go beyond simple content generation.

Naomi King offered an insightful case study from her own experience creating a hiring task for intern developers. To ensure candidates could not just copy and paste a solution, she deliberately engineered a flaw in the assignment: "I made this fake database that intern candidates would pull tickets from, and 1 in 10 times they pulled tickets, it would be a corrupted ticket format." An AI-generated solution would likely fail at this unexpected hurdle because it would create a general solution that does not account for error handling. In contrast, she explained, a human programmer "would have to run it a bunch of times... and they'd realize, 'oh, this is really weird.'" **In other words, this type of task design does not necessarily forbid AI, but it makes it insufficient on its own, thereby rewarding deeper thinking and engagement.** This aligns with her broader recommendation for educators to focus on higher-order skills like "system design because AI can't really do that right yet."

Tasks like Naomi King's hiring challenge also point toward a broader principle:

Rather than setting predetermined solutions, effective assessments in the age of AI are often about real-time problem-solving and adaptability. As discussed in the following, this naturally leads toward more interactive and 'on-the-fly' assessment formats where learners must demonstrate their thinking process and not just their final answers.

## **3. The Resurgence of Dialogues and Viva Voce**

In parallel with redesigning written tasks, educators are rediscovering the power of dialogues and oral check-ins as a tool for assessment. As Tal Havivi pointed out, given that polished text can now be generated by AI instantly, "What are the ways that we can measure [genuine learning] now and how should we measure that differently?" In our conversations, we notice that many educators are turning to assessing learners' ability to articulate their reasoning in real-time.

Dr Anne-Marie Stanton-Ife noted that the viva voce, or oral exam, is becoming "a tool much more widely used across departments, not only to support oracy skills but also for checking for understanding." This shift is happening even in fields that have long relied on written tests. She finds it "quite revealing" that chemistry departments, for example, are now embracing "oral questioning and recording oral responses, rather than 'sit down, here is your 45-minute test.'" These methods force learners to move beyond recall and demonstrate ownership of their knowledge; they must be able to explain the why and how behind their work, instead of just producing the correct answer.

This pivot toward dialogue, Stanton-Ife observed, fundamentally changes the classroom assessment atmosphere. “The expectations of that kind of participation have become slightly higher stakes... [Learners] know they’re going to have to be able to explain themselves rather than just write down the answer.” As a result, they are “coming around to seeing that as a viable form of assessment.” Of course, this interactive validation is not limited to summative exams; it is also powerful when woven into the daily learning process and formative assessments. Steph Chambers, for instance, noted that some teachers would ask “face-to-face questions about [students’] essays” when it comes to writing assignments. Rachel Bowen also explained that, in subjects like maths, “it’s not just looking at the final answer; it’s making sure that students can show their working in something like a ‘working book’ so we can check that the process is happening.” She also adds that teachers can have informal conversations with students “while they are working—even when they’re using AI... and spend about five minutes with each student just to make sure that understanding is there, giving them ‘a little bit of a push in the right direction.’”

This approach operates on a simple, powerful recognition: **You can’t prevent the use of AI, and that’s not what really matters.** The ultimate goal is for learners to genuinely understand their work. If a learner can coherently explain their reasoning when questioned and demonstrate ownership of the thinking that produced it, then they have met the learning objective, regardless of what tools they used along the way.

#### **4. Blend Modalities: Combine AI-Assisted Process with Unaided Production**

Other than conducting live conversations, educators are designing blended assessments that separate the AI-assisted learning processes from the unaided final product, allowing learners to leverage AI for ideation and exploration while requiring them to demonstrate what they have internalized. Ha Thi Hai Nguyen provided a clear example of this method in practice: She allows her students to use AI for brainstorming and then asks them to “turn off all the electric devices and... write by hand.” This strategy provides the scaffolding and space for learners to experiment and learn with AI to lower the barrier to starting a task, but maintains the cognitive demand of synthesis and recall. The goal in these blended modalities remains the same: To assess genuine understanding of the material. As one of Conrado Torres’s teachers put it, “You can use AI, you can use everything you want—I’m going to realize if you don’t really make the connections personally.” The assessment, therefore, becomes a measure of that internalized understanding, not just the ability to prompt an AI.

**Ultimately, redesigning assessment should not be a defensive tactic against AI, but a proactive strategy to cultivate more authentic and durable learning by rewarding the process, valuing intellectual struggle, and designing tasks that demand critical human engagement.** In order to do so, schools must be intentional in creating an environment that recognizes AI as a resourceful tool for learning instead of a shortcut around it, enabling educators to confidently assess genuine understanding and motivate learners

to develop the resilient, adaptive skills essential for a future where they will work alongside, not in opposition to, intelligent technologies.

### **Empowering Educators as AI Learning Architects**

From a teacher professional development perspective, it is important to think beyond efficiency and consider what constitutes effective and high-quality teaching—the former simply ‘makes teaching easier’, but the latter ‘makes teaching better’. Dr Maria Ruiz-Primo emphasized that effective AI use requires deep pedagogical reflection: “From the teacher’s side is to use AI in a way that helps them to reflect on their own knowledge and what they can understand better by using AI.” She further highlighted a critical fault line in AI adoption:

**[Using] AI is not about asking questions and getting responses. It’s not about asking: Give me five multiple-choice test questions that have these characteristics... That won’t help if [educators] don’t reflect on the evidence of learning that they are trying to gather with those five questions.**

As such, **rather than being seduced by techno-solutionist capabilities by approaching AI with surface-level requests for lesson plans or assessments, educators should leverage AI as a thinking partner while maintaining agency over their educational decisions.**

Dr Guillermo Solano-Flores extended this perspective by broadening the discourse to our relationship with knowledge: “If we don’t want to become victims as individu-

als of artificial intelligence and we want to grow, we have to see ourselves not as entities who know but entities who manage knowledge.” This shift from knowledge holders to knowledge managers directly resonates with Dr Maria Ruiz-Primo’s distinction between surface-level and reflective AI use, and represents a reimagining of professional identity in education: Since AI can instantly access and synthesize vast amounts of information, teachers can no longer rely primarily on their role as repositories of subject matter knowledge; instead, their professional value lies in their ability to critically curate, evaluate, and adapt knowledge for specific learners and learning contexts.

Ha Thi Hai Nguyen, for instance, exemplified this approach of knowledge management:

**Whenever I plan a lesson and I work with a co-pilot... When we understand the methodology, then we can assign [AI] to do this, and we can consider whether it is actually innovative or not. And I do appreciate the way they create different types of exercises for my students because when they create the exercises, I find it brand new—nothing in any books.**

Essentially, she maintains pedagogical control by first architecting the methodological and knowledge frameworks—understanding the ‘why’ behind her teaching before delegating the ‘what’ to AI. Her appreciation for AI’s capacity to generate new exercises is thus grounded within (rather than detached from) her educational judgement.

# Good Practices: Assigning AI Pedagogical Roles to Support Learning

When it comes to using some of the more widely available or accessible tools, what is crucial from an educators' perspective is actively experimenting, iterating, and interacting with these tools critically. Drawing from Mollick and Mollick's (2023) framework, below are some strategies and approaches that educators can directly implement to support learners.

## Deploy AI in 7 different pedagogical roles:

- 1. AI as Mentor** – Use AI to provide frequent, balanced feedback on learners' work. Design prompts that ask the AI to identify strengths and areas for improvement, while encouraging students to critically evaluate, challenge, and act on this feedback.
- 2. AI as Tutor** – Prompt AI to offer personalized instruction to learners by assessing their prior knowledge, adapting explanations to their learning levels, and guiding them through concepts via questioning rather than simply providing answers.
- 3. AI as Coach** – Facilitate metacognitive reflection by having AI guide learners through structured thinking processes about their learning, project planning, and problem-solving approaches.
- 4. AI as Teammate** – Use AI to help teams understand member strengths, play devil's advocate on decisions, or provide alternative perspectives on group challenges.
- 5. AI as Student** – Strengthen understanding by having learners "teach" the AI concepts, then evaluate and correct the AI's explanations or work to reinforce their own mastery.
- 6. AI as Simulator** – Create opportunities for deliberate practice where learners can apply skills and receive immediate feedback in simulated scenarios.
- 7. AI as Tool** – Extend learners' capabilities by using AI to accomplish routine tasks more efficiently.

However, it is important to note that each approach also carries pedagogical risks that educators must proactively take into consideration. For instance, AI may provide misleading feedback or confabulated information, lack awareness of individual student learning styles or contexts, and risk fostering over-dependence that undermines critical thinking. As such, as educators and learners are experimenting with these approaches, an important goal is to also learn to critically and responsibly evaluate, verify, and build upon AI outputs.

## Learning the Value of Learning

A growing body of research and discourse have emphasized balancing the design and use of technologies with fostering students' critical thinking, metacognitive skills, and 'productive struggle'—the process of engaging in challenging tasks that require effort, deliberate difficulties, critical thinking, and perseverance that enhance the metacognition crucial for deeper learning (Griffin & James, 2025; Young et al., 2023). However, before we address the practical dimensions of AI adoption, we need to return to the first principle: ensuring that learners understand what 'learning' actually is, what it means, and why it matters. Dr Maria Ruiz-Primo highlighted how crucial it is to have learners learn the value of 'learning':

**Being reflective and metacognitive is so critical... If students do not understand that they are also the agents of their own learning, there is so little that you can do, right? Because they will always blame you for what they did learn or did not learn or whatever.**

Developing this metacognitive awareness enables them to take genuine responsibility for their own intellectual growth and to value learning for its intrinsic worth, rather than merely for external rewards like grades or performance. Ha Thi Hai Nguyen also connected this agency directly to the realities of assessments:

**We really need to raise the responsibility of the students, because in the real test, they do not have any assistants at all. In order to gain the better band, it is their responsibility to be truthful with themselves.**

More broadly, though, the cultivation of learners' responsibility must be situated in a wider debate about whether internal knowledge remains important at all in an age where information can be easily accessed through search engines and AI tools. In today's age, we can outsource information storage to the internet and digital systems as long as we retain the 'biological pointers'—the memory traces that direct us where to find what we need, rather than the information itself (Skulmowski, 2023). However, research suggests that this creates **not only an emotional alienation from their own learning process, but also, more concerningly, an illusion of competence**: Learners may produce high-quality outputs or feel that they have learned information, but in reality become trapped in a vicious cycle—where reduced mental effort weakens working memory and impairs schematic formation, which in turn increases reliance on external tools, and further prevents the development of the cognitive architecture necessary for deeper knowledge and expertise (Oakley et al., 2025).

What we are really talking about when we speak of 'learning the value of learning' is the development of self-regulation, which consists of three interwoven dimensions that empower learners to take ownership of their learning journey (Luckin, 2025b):

**1. Cognitive elements:** The intentional implementation of learning processes and strategies.

**2. Metacognitive elements:** The ability to reflect on the process of learning (rather than just the content) via analyzing learning tasks, setting goals, monitoring progress, and adapting strategies.

**3. Motivational elements:** The persistence, confidence, and emotional resilience needed to sustain engagement through challenging learning experiences in order to arrive at genuine understanding.

In the context of AI use in education, these three dimensions highlight that learners should **learn to leverage AI as a thinking partner in intentional ways to approach tasks effectively, recognize when AI is scaffolding or hindering their own thinking, and maintain intrinsic motivation despite the ease of generating answers with AI.** Furthermore, beyond individual

self-regulation (where learners manage their own learning processes), the social forms of learning—such as co-regulation (where peers support each other's learning) and socially shared regulation (coordinated group learning)—are equally important (Hadwin et al., 2018). Not just collaboration amongst humans, but also human-AI collaboration, can facilitate synergistic work that supports cognitive and metacognitive activities within collaborative learning contexts (Edwards et al., 2024; Järvelä et al., 2023).

## Good Practices: Designing AI Prompts to Support Learner Agency

On top of the 7 pedagogical roles as described previously, Mollick and Mollick (2023) also identified components to be included in effective AI prompt engineering for educational purposes. In the following, we adapt these components specifically to support the cognitive, metacognitive, and motivational dimensions of self-regulated, co-regulated, and socially shared regulated learning.

### **1. Role and Goal – Clearly specify who the AI should be and articulate what you want it to accomplish.**

Example: “You are a thinking partner who helps students develop their own understanding of [topic]. Your goal is to guide students to construct knowledge rather than provide direct answers.”

### **2. Step-by-Step Instructions – Break down interactions into manageable sequences; promote regulation by building in reflection points, goal-setting opportunities, and strategy monitoring.**

Example: “First, ask learners to set their learning goals. Then guide them to choose appropriate strategies. After each step, prompt them to evaluate their progress and adjust their approach.”

**3. Pedagogy – Embed pedagogical approaches and learning theories that emphasize student agency, social and metacognitive development, and productive struggle. Identify common misconceptions and include strategies to address them through active learning methods.**

Example: “Adopt an inquiry-based approach by guiding learners to come up with their own questions, then scaffold them to devise plans to investigate their inquiries methodically. For group work, offer questions that team members can discuss together to deepen their collective understanding.”

**4. Constraints – Set clear limits on what the AI should and should not do in order to prevent over-reliance and promote genuine learning.**

Example: “Do not directly provide solutions; instead, push the learner explain their reasoning in their own words. For group work, encourage learners to discuss and validate AI suggestions together, do not proceed until you get a response.”

**5. Personalization – Adapt to individual learning contexts by encouraging students to gauge their prior knowledge and level of understanding, and recognize their learning patterns and preferences.**

Example: “Have students to rate their confidence level on the topic and identify specific areas where they feel uncertain. Then ask them to reflect on what learning strategies work best for them and why. Use their responses to tailor your instructional approach accordingly.”

By building in mechanisms for self-regulated learning, learners can document their AI interactions, reflect on the effectiveness of different approaches, and evaluate how AI use supports or hinders their learning goals. As demonstrated, in collaborative contexts, this extends to co-regulation and socially shared regulation, where learners help each other monitor AI use, maintain accountability for genuine learning, and collectively decide how to integrate AI in ways that enhance knowledge co-construction.

For more prompts for educator use, you may also visit:

<https://www.aiforeducation.io/prompt-library>

Finally, all of this comes down to: How do we want to prepare learners for the future? Are our approaches empowering them to learn for their own sake, or foster dependency? In an age no longer characterized by information scarcity and predictable career paths, but rather abundance, ubiquity, and constant change, greater value is placed on higher-order thinking skills and adaptability. As AI systems increasingly handle routine cognitive work, this trend will only continue to hold true.

Yet, the paradox is that while AI renders information instantly accessible, foundational knowledge becomes more critical, not less. In order to think critically,

problem-solve creatively, and evaluate or leverage AI-generated information meaningfully, we need subject matter expertise (Mollick, 2024). The future belongs not to those with static knowledge but to those who can continuously learn and adapt, yet this capacity itself requires the foundation that enables learners to identify what they need to learn next and integrate it meaningfully (Luckin, 2025b). Learning the value of learning thus becomes essential preparation for a world where human insight and judgment grow more valuable precisely because they enable us to work effectively alongside AI tools.

## Conclusion

As we have emphasized throughout this research, responsibly integrating AI into education is not just a technical problem to be solved, but a human endeavor that requires educators, learners, and policymakers to move beyond piecemeal solutions and adopt a coherent, human-centered strategy. The three interdependent pillars underpinning this research highlight how we need deliberate and well-intentioned steps that form a virtuous cycle: It begins with mindset, shifting a learning community's initial fear and uncertainty into confidence built upon transparency and critical literacy. This confidence, however, must be anchored by robust mechanisms of institutional support—from collaborative vetting processes to peer-led professional development—that provide the practical guardrails for responsible adoption, lest optimism become recklessness. Finally, both confidence and support must serve a higher purpose, addressing the core motivation of education: By redesigning assessment to reward process over product and empowering educators as learning architects, we can inspire learners' genuine desire for deep understanding in an age where superficial answers are ubiquitous. **These pillars are by no means separate checklists, but a dynamic interplay forming the basis of**

**sustainable ecosystems for responsible and effective AI adoption in education—the confidence to engage, the support to do so safely, and the motivation to do so with purpose.**

The discourse surrounding AI in schools must transcend the tools themselves. It is a conversation about our fundamental values and what it means to learn, grow, and flourish. Yes, AI can generate text, solve equations, and optimize workflows, but it cannot replicate or replace the authentic dialogues, the messy collaborative discoveries, the productive struggle that builds resilience, or the mentorship founded upon our capacity for care and intentionality that lies at the heart of meaningful education. **Our task, then, is not so much to simply integrate a new technology, but to leverage it in service of our most enduring goals: cultivating wisdom, empathy, and the shared human capacity for making meaning.** This is a task beyond simply avoiding the pitfalls of AI, but using it as a mirror and a chisel—one that reflects, clarifies, and shapes what is, and always has been, uniquely and irreplaceably human about learning.

**“The path forward requires courage and persistence: the courage to maintain human relationships in an age of artificial ones, to value emotional development alongside academic achievement, and the persistence to insist that technology serve rather than subvert our deepest educational values. It requires wisdom to discern when AI enhances learning and when it diminishes it, when to provide support and when to allow struggle, when to connect digitally and when to insist on face-to-face presence. Most importantly, it requires love, the love that motivated Tomasello’s cooperative creatures to share knowledge across generations, that drives teachers to nurture their students’ growth, and that inspires learners to persist despite challenges. No AI, no matter how sophisticated, can replicate this fundamentally human capacity for care.”**

**—Professor Rose Luckin, 2025  
Professor of Learner Centred Design at UCL Knowledge Lab,  
Founder & CEO of EDUCATE Ventures**

Research by Xinman Liu, Joseph Lin, and Daniel Emmerson.  
For citations, please cite ‘Goodnotes, 2025’

### **Transparency Statement**

All content, research questions, methodology, stakeholder interviews, and core arguments were conceived, conducted, and written by human researchers. AI tools, specifically Goodnotes AI, Google Gemini, Notebook LM, Claude Sonnet 4, Perplexity, Elicit, and Research Rabbit, were used for literature review, interview synthesis, and writing refinement. Any AI-generated content, when used, was critically reviewed and revised by the authors. The authors maintain full responsibility for the accuracy and integrity of all content.

# Acknowledgements

**Dr Andy Kemp**

Principal, The National Mathematics  
and Science College

**Dr Anne-Marie Stanton-Ife**

Deputy Head Academic,  
Headington Rye School

**Chi-Hung Ha**

Chairman, FlippEducators@HK;  
IT Panel Head,  
True Light Middle School of Hong Kong

**Conrado Torres**

Student, Torcuato Di Tella University

**Ha Thi Hai Nguyen**

Dr of Education Candidate and Lecturer at  
Vietnam National University in Hanoi

**John Jones**

Assistant Head,  
Royal Grammar Schools Worcester

**Kate Atherton**

Associate Assistant Headteacher at St  
John the Baptist School

**Naomi King**

Student, Queensland University  
of Technology;  
Customer Support Operations Engineer,  
Goodnotes

**Nathan Nagaiah**

Lead of UK Centre for AI and Strategic  
Lead for Artificial Intelligence, Data and  
Partnerships at London Borough of  
Newham

**Rachel Bowen**

Assistant Head at Headington  
Rye School in Oxford

**Dr Maria Ruiz-Primo**

Associate Professor at Stanford  
Graduate School of Education

**Selin O.**

Student, Robert College

**Dr Guillermo Solano-Flores**

Professor at the Stanford Graduate  
School of Education

**Steph Chambers**

Digital Innovation Lead at ACS Egham

**Tal Havivi**

Managing Director for Research &  
Development at ISTE

# References

- AI4K12.org. (2021). The State of K-12 AI Education in Your State Workshop: Executive Summary. AI4K12 Initiative. <https://ai4k12.org/wp-content/uploads/2021/08/Executive-Summary-AI-Education-in-Your-State-v1.0.docx.pdf>
- AI for Education. (2025). State AI Guidance for Education. AI for Education. <https://www.aiforeducation.io/ai-resources/state-ai-guidance>
- Bao, A., & Zeng, Y. (2025). AI disclosure, moral shame, and the punishment of honesty. *Accountability in Research*, 1–14. <https://doi.org/10.1080/08989621.2025.2542197>
- Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., & Sailer, M. (2025). Looking Beyond the Hype: Understanding the Effects of AI on Learning. *Educational Psychology Review*, 37(2), 45. <https://doi.org/10.1007/s10648-025-10020-8>
- Baxter, D. (2024). Teachers using ChatGPT – alongside a guide to support them to use it.... EEF. <https://educationendowmentfoundation.org.uk/news/teachers-using-chatgpt-alongside-a-guide-to-support-them-to-use-it-effectively-can-cut-lesson-planning-time-by-over-30-per-cent>
- Bellan, R. (2025). AI sycophancy isn't just a quirk, experts consider it a "dark pattern" to turn users into profit. *TechCrunch*. <https://techcrunch.com/2025/08/25/ai-sycophancy-isnt-just-a-quirk-experts-consider-it-a-dark-pattern-to-turn-users-into-profit/>
- Berendt, B., Littlejohn, A., & Blakemore, M. (2020). AI in education: Learner choice and fundamental rights. *Learning, Media and Technology*, 45(3), 312–324. <https://doi.org/10.1080/17439884.2020.1786399>
- Bernhardt, M. (2025). Bridging the AI Competency Gap: From Literacy to Fluency in L&D. *Learning Guild*. <https://www.learningguild.com/articles/bridging-the-ai-competency-gap-from-literacy-to-fluency-in-l&d>
- Blackmon, S. J. (2023). Student Privacy and Data Literacy: An Educational Opportunity. *Change: The Magazine of Higher Learning*, 55(6), 21–28. <https://doi.org/10.1080/00091383.2023.2263189>
- Breazeal, C., & Lee, I. (2022). The DAILY Curriculum for Middle School Students. MIT RAISE. <https://raise.mit.edu/daily/>
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Campbell, S., Nguyen, R., Bonsignore, E., Carter, B., & Neubauer, C. (2023). Defining and Modeling AI Technical Fluency for Effective Human Machine Interaction. *Human Factors in Robots, Drones and Unmanned Systems*, 93(93). <https://doi.org/10.54941/ahfe1003743>
- Chan, C. K. Y., & Colloton, T. (2024). AI Literacy. In *Generative AI in Higher Education: The ChatGPT Effect*. Routledge. <https://doi.org/10.4324/9781003459026>
- Chen, K., Zhou, X., Lin, Y., Feng, S., Shen, L., & Wu, P. (2025). A Survey on Privacy Risks and Protection in Large Language Models (No. arXiv:2505.01976; Version 1). *arXiv*. <https://doi.org/10.48550/arXiv.2505.01976>
- Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501–6506 (1998).

CNIL. (2022). Online age verification: Balancing privacy and the protection of minors. Commission nationale de l'information et des libertés (CNIL). <https://www.cnil.fr/en/online-age-verification-balancing-privacy-and-protection-minors>

Dakan, R. & Feller, J. (2025). The AI fluency framework: Four interconnected competencies necessary to ensure our interactions with AI are effective, efficient, ethical and safe [Framework]. Anthropic.

Duffy, C. (2024). 'There are no guardrails.' This mom believes an AI chatbot is responsible for her son's suicide. CNN. <https://www.cnn.com/2024/10/30/tech/teen-suicide-character-ai-lawsuit>

Du, H., Niyato, D., Kang, J., Xiong, Z., Zhang, P., Cui, S., Shen, X., Mao, S., Han, Z., Jamalipour, A., Poor, H. V., & Kim, D. I. (2024). The Age of Generative AI and AI-Generated Everything. *IEEE Network*, 38(6), 501–512. <https://doi.org/10.1109/MNET.2024.3422241>

Do, Y. (2025). Beyond Privacy: Regulating ChatGPT for Young Adults in Educational Contexts. *New York University Journal of Intellectual Property & Entertainment Law (JIPEL)*, 14(2). <https://jipel.law.nyu.edu/beyond-privacy-regulating-chatgpt-for-young-adults-in-educational-contexts/>

Edwards, J., Nguyen, A., Lämsä, J., Sobocinski, M., Whitehead, R., Dang, B., Roberts, A.-S., & Järvelä, S. (2025). Human-AI collaboration: Designing artificial agents to facilitate socially shared regulation among learners. *British Journal of Educational Technology*, 56(2), 712–733. <https://doi.org/10.1111/bjet.13534>

European Parliament and Council (2018). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119/1. <http://data.europa.eu/eli/reg/2024/1689/oj/eng>

[reg/2024/1689/oj/eng](http://data.europa.eu/eli/reg/2024/1689/oj/eng)

European Parliament and Council (2024). Regulation (EU) 2024/1689 Laying down Harmonised Rules on Artificial Intelligence, OJ L 213/1. <http://data.europa.eu/eli/reg/2024/1689/oj/eng>

Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g (1974).

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489–530. <https://doi.org/10.1111/bjet.13544>

Forland, S. (2025). Exploring Privacy-Preserving Age Verification: A Close Look at Zero-Knowledge Proofs. *New America*. <http://newamerica.org/oti/briefs/exploring-privacy-preserving-age-verification/>

Fire, M., Elbazis, Y., Wasenstein, A., & Rokach, L. (2025). Dark LLMs: The Growing Threat of Unaligned AI Models (No. arXiv:2505.10066). arXiv. <https://doi.org/10.48550/arXiv.2505.10066>

Goodnotes (2023). Generative AI in Schools: Evolving Coursework, Homework and Assessment in the Age of AI. <https://www.goodnotes.com/research/generative-ai-in-schools>

Goodnotes (2024). Social and Emotional Impact of Generative AI in Schools. <https://www.goodnotes.com/research/social-emotional-impact-generative-ai-schools>

Griffin, H., & James, T. (2025). The temptation of the silver platter: Preventing the sidestep and normalising the productive struggle. *Intelligent Technologies in Education*. <https://doi.org/10.70770/tbb3db47>

Gulson, K. N., Sellar, S., & Webb, P. T. (2022).

- Algorithms of Education: How Datafication and Artificial Intelligence Shape Policy. University of Minnesota Press. <https://www.jstor.org/stable/10.5749/j.ctv2fzkpxp>
- Hadwin, A., Järvelä, S., & Miller, M. (2018). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In *Handbook of self-regulation of learning and performance*, 2nd ed (pp. 83–106). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9781315697048-6>
- Holmes, W., Mouta, A., Hillman, V., Schiff, D., Laak, K.-J., Atenas, J., Bardone, E., Lochead, K., Gonsales, P., Havemann, L., Seon, J., Go, B., Schreurs, B., Zhgenti, S., Lee, K., Bali, M., Bialik, M., Medina-Gual, L., Adhichandra, I., Anwar, T., Arantes, J., Baten, D., Bermudez Macias, E., Bowditch, I., Cesaroni, V., Conrad, K., Cox, A., Eitner, A., Gibson, P., Haniya, S., Hau, D., Hogan, M., Janakievskia, G., Jaris, J., Johri, A., Karademir, N., Komljenovic, J., Kurian, N., LaRaia, M., Leaton Gray, S., Liu, X., Mavrikis, M., Monett, D., Mor, Y., Nalyvaiko, O., Neri, P., Pajula, L., Parchimowicz, M., Prinsloo, P., Qadir, J., Ruttkamp-Bloem, E., Sabourin, B., Salzmänn, A., Satia, A., Schleiß, J., Şenocak, D., Shiohira, K., Soudi, M., Tali Otmani, F., Tapp Jaksa, A., Westermann, W., Williamson, B., & Yeo, B. (2025). Critical studies of artificial intelligence and education: Putting a stake in the ground (SSRN Working Paper No. 5391793). <https://dx.doi.org/10.2139/ssrn.5391793>
- Holthouse, R., Owens, S., & Bhunia, S. (2025). The 23andMe Data Breach: Analyzing Credential Stuffing Attacks, Security Vulnerabilities, and Mitigation Strategies (No. arXiv:2502.04303). arXiv. <https://doi.org/10.48550/arXiv.2502.04303>
- Horwitz, J. (2025). Meta's AI rules have let bots hold 'sensual' chats with children. Reuters. <https://www.reuters.com/investigates/special-report/meta-ai-chat-bot-guidelines/>
- Howard, S. K., Swist, T., Gasevic, D., Bartimote, K., Knight, S., Gulson, K., Apps, T., Peloché, J., Hutchinson, N., & Selwyn, N. (2022). Educational data journeys: Where are we going, what are we taking and making for AI? *Computers and Education: Artificial Intelligence*, 3, 100073. <https://doi.org/10.1016/j.caeai.2022.100073>
- Huang, L. (2023). Ethics of Artificial Intelligence in Education: Student Privacy and Data Protection. *Science Insights Education Frontiers*, 16(2), 2577–2587. <https://doi.org/10.15354/sief.23.re202>
- ICO. (2024, November 27). What is a DPIA? ICO. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/data-protection-impact-assessments-dpias/what-is-a-dpia/>
- Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*, 54(5), 1057–1076. <https://doi.org/10.1111/bjet.13325>
- Kaithathara, S., & Jose, A. (2025). Empowering Child Safety Using AI and ML to Address Online Risks. In R. Rawat, S. K. Sarangi, A. S. Arun Raj, J. O. Richmond, & P. Bhardwaj (Eds.), *Child protection laws and crime in the digital era*. IGI Global Scientific Publishing.
- Keane, W. (2025). *Animals, robots, gods: Adventures in the moral imagination*. Princeton University Press.
- Kelso, E., Soneji, A., Rahaman, S., Shoshitaishvili, Y., & Hasan, R. (2024). Trust, Because You Can't Verify: Privacy and Security Hurdles in Education Technology Acquisition Practices. *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, 1656–1670. <https://doi.org/10.1145/3658644.3690353>
- Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V.,

Braunstein, I., & Maes, P. (2025). Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task (No. arXiv:2506.08872). arXiv. <https://doi.org/10.48550/arXiv.2506.08872>

Lee, H.-P. (Hank), Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 1-22. <https://doi.org/10.1145/3706598.3713778>

Luckin, R. (2025a). What the research says about: Emotions as the Foundation for Self-Directed Learning in an AI-Enhanced World. Educate Ventures Research: What the Research Says. <https://www.educateventures.com/what-the-research-says>

Luckin, R. (2025b). What the research says about: Self-Directed and Self-Regulated Learning in an AI-Driven World. Educate Ventures Research: What the Research Says. <https://www.educateventures.com/what-the-research-says>

Marlowe, C. (2025). Children's Online Privacy in 2025: The Amended COPPA Rule. Loeb & Loeb LLP. <https://www.loeb.com/en/insights/publications/2025/05/childrens-online-privacy-in-2025-the-amended-coppa-rule>

Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO. <https://www.unesco.org/en/articles/ai-competency-framework-teachers>

Mollick, E. (2024). Co-intelligence: Living and working with AI. Portfolio/Penguin.

Mollick, E. (2025). Making AI Work: Leadership, Lab, and Crowd. One Useful Thing. <https://www.oneusefulthing.org/p/making-ai-work-leadership-lab-and>

Mollick, E. R., & Mollick, L. (2023). Assigning AI: Seven Approaches for Students, with Prompts (SSRN Scholarly Paper No. 4475995). Social Science Research Network. <https://doi.org/10.2139/ssrn.4475995>

Meyerhofer, K. (2021). UW-Madison Renews Contract With Controversial Exam Software. GovTech. <https://www.govtech.com/education/higher-ed/uw-madison-renews-contract-with-controversial-exam-software>

Mollick, E. (2024). Co-intelligence: Living and working with AI. Portfolio/Penguin.

Moore, J., Grabb, D., Agnew, W., Klyman, K., Chancellor, S., Ong, D. C., & Haber, N. (2025). Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers. Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency, 599-627. <https://doi.org/10.1145/3715275.3732039>

Norden, J. (2025). Safeguarding leads call for clarity on publication of KCSIE 2025 [Tes]. Tes Magazine. <https://www.tes.com/magazine/news/general/delayed-kcsie-safeguarding-guidance-causes-anxiety-for-schools>

Oakley, B., Johnston, M., Chen, K., Jung, E., & Sejnowski, T. (2025). The Memory Paradox: Why Our Brains Need Knowledge in an Age of AI. OSF. [https://doi.org/10.31234/osf.io/3xye5\\_v2](https://doi.org/10.31234/osf.io/3xye5_v2)

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. Educational Psychology Review, 18(4), 315-341. <https://doi.org/10.1007/s10648-006-9029-9>

Peltz, J., & Street, A. C. (2020). Artificial Intelligence and Ethical Dilemmas Involving Privacy. In Artificial Intelligence and Global Security (pp. 95-120). Emerald Publishing Limited.

<https://doi.org/10.1108/978-1-78973-811-720201006>

Piri, C. (2024). Data Privacy in the age of LLM-based services in Education: Current Challenges, Improvement Guidelines and Future Directions [University of Jyväskylä]. <https://urn.fi/URN:NBN:fi:ju-202409306177>

Raffaghelli, J. E., & Stewart, B. (2020). Centering complexity in 'educators' data literacy' to support future practices in faculty development: A systematic review of the literature. *Teaching in Higher Education*, 25(4), 435–455. <https://doi.org/10.1080/13562517.2019.1696301>

Reich, J. (2020). *Failure to disrupt: Why technology alone can't transform education*. Harvard University Press.

Reich, J. (2023). *Iterate: The secret to innovation in schools*. Jossey-Bass.

Saarela, M., Gunasekara, S., & Karimov, A. (2025). The EU AI Act: Implications for Ethical AI in Education. In S. Chatterjee, J. vom Brocke, & R. Anderson (Eds.), *Local Solutions for Global Challenges* (pp. 36–50). Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-93979-2\\_3](https://doi.org/10.1007/978-3-031-93979-2_3)

Soderstrom, N. C., & Bjork, R. A. (2015). Learning Versus Performance: An Integrative Review. *Perspectives on Psychological Science*, 10(2), 176–199. <https://doi.org/10.1177/1745691615569000>

Sood, A. K., & Zeadally, S. (2025). The Unprecedented Surge in Generative AI: Empirical Analysis of Trusted and Malicious Large Language Models (LLMs). *IEEE Technology and Society Magazine*, 1–11. <https://doi.org/10.1109/MTS.2025.3582667>

Shen, L., Chen, I., & Su, A. (2017). Cybersecurity and Data Breaches at Schools. In *Advances in Information Security, Privacy, and Ethics* (pp. 144–174). IGI Global. <https://doi.org/10.4018/978-1-5225-1941-6.ch007>

Skulmowski, A. (2023). The Cognitive Architecture of Digital Externalization. *Educational Psychology Review*, 35(4), 101. <https://doi.org/10.1007/s10648-023-09818-1>

Staab, R., Vero, M., Balunović, M., & Vechev, M. (2024). Beyond Memorization: Violating Privacy Via Inference with Large Language Models (No. arXiv:2310.07298; Version 2). arXiv. <https://doi.org/10.48550/arXiv.2310.0729>

SWGfL. (n.d.). SWGfL Data Protection Guidance—Part 9—DPIA and DPO. Retrieved July 11, 2025, from <https://swgfl.org.uk/resources/gdpr-guidance-for-schools-and-colleges/part-9-dpia-and-dpo/>

The White House. (2025, April 23). Advancing Artificial Intelligence Education for American Youth. <https://www.whitehouse.gov/presidential-actions/2025/04/advancing-artificial-intelligence-education-for-american-youth/>

UNESCO. (2021). *AI and education: Guidance for policy-makers*. UNESCO. <https://doi.org/10.54675/pcsp7350>

Viberg, O., Cukurova, M., Feldman-Maggor, Y., Alexandron, G., Shirai, S., Kanemune, S., Wasson, B., Tømte, C., Spikol, D., Milrad, M., Coelho, R., & Kizilcec, R. F. (2024). What Explains Teachers' Trust of AI in Education across Six Countries? (No. arXiv:2312.01627). arXiv. <https://doi.org/10.48550/arXiv.2312.01627>

Wijaya, T. T., Yu, Q., Cao, Y., He, Y., & Leung, F. K. S. (2024). Latent Profile Analysis of AI Literacy and Trust in Mathematics Teachers and Their Relations with AI Dependency and 21st-Century Skills. *Behavioral Sciences*, 14(11), Article 11. <https://doi.org/10.3390/bs1411008>

Wineburg, S. (2025). School's In: Digital literacy in the AI era (Part 1) (D. Pope & D. Schwartz, Interviewers) [Interview].

<https://ed.stanford.edu/news/digital-literacy-ai-era-part-1>

Young, J. R., Bevan, D., & Sanders, M. (2023). How Productive is the Productive Struggle? Lessons Learned from a Scoping Review. *International Journal of Education in Mathematics, Science and Technology*, 12(2), 470–495. <https://doi.org/10.46328/ijemst.3364>

Zhong, V., McGregor, S., & Greenstadt, R. (2023). “I’m going to trust this until it burns me” Parents’ Privacy Concerns and Delegation of Trust in K-8 Educational Technology. 5073–5090. <https://www.usenix.org/conference/usenixsecurity23/presentation/zhong>