AI Explained · Why We Fear AI, Part 3

AI in the Classroom

By Dr Peter Fields

THE 2-MINUTE SUMMARY

A polished essay tells a teacher less if a chatbot did much of the thinking. The same technology can give a pupil a useful hint at the moment they are stuck.

In a trial in secondary-school maths, a standard chatbot improved practice marks but reduced scores when pupils later worked unaided. A tutor designed with teachers' input produced a different result.

Decide which skill the pupil is meant to practise. Ask for evidence of that skill, and take care before accusing someone of handing in work that is not their own.

The essay that arrived on time

A teacher has thirty essays to mark. One is clear and well organised, even though its author struggled to make an argument in class. They may have worked hard on it. They may have used a chatbot to get past a difficult paragraph, or to write the whole thing. The finished essay leaves the teacher with a problem. What can the pupil now do on their own?

A calculator can do the arithmetic after a pupil has decided which calculation to make. A chatbot can also interpret the question, choose the method and explain the answer. That breadth matters when any one of those steps is the skill the lesson is intended to teach. Handing off a task can even be sensible. A student writing a history essay might use an accessibility tool to transcribe their ideas, then work hard on the quality of the argument.

Which difficulty is worth keeping?

Bjork and Bjork use the term 'desirable difficulties' for conditions that can make practice feel harder while improving what a learner can recall or apply later. Recalling an idea before looking it up is one example. A difficulty earns its place when it helps develop a skill. A barrier to getting started may deserve to be removed. Research on cognitive offloading gives a useful way to think about the choice: what work are you handing to the tool, and what will you do with the attention it frees?[1,2]

Geek Box · What pupils could do without the tool

In a trial involving nearly 1,000 Turkish secondary-school maths pupils, Bastani and colleagues found practice scores 48% higher with a standard GPT-4 interface and 127% higher with a version using teacher-supplied guidance. On an unaided assessment, the standard chatbot group scored 17% below the control group. The guided group's unaided scores did not differ significantly from the control group's. Those results concern the particular tools, lessons and assessments in the trial.[3]

In a university experiment, Contractor and Reyes found that AI access raised immediate scores on an unaided knowledge test by 0.27 standard deviations, a measure of the size of the difference. Knowledge

gains persisted a week later. Students who used AI for explanations showed stronger delayed writing gains than those who used it largely to generate text. The short-run gains in writing quality for the latter group faded once the tool was removed. The researchers measured knowledge and writing separately, which is precisely what made the result useful.[4]

A trial by Kestin and colleagues adds a reason to take tutoring seriously. Undergraduates studying selected physics lessons learnt more in less time with a tutor designed for those lessons than in the active-learning class to which it was compared. That is evidence for a particular design. A teacher still has to decide what evidence of learning matters and how to respond when a pupil is struggling.[5]

When a teacher suspects AI use

A detector is a poor basis on its own for accusing someone of cheating. Liang and colleagues found that the tools they tested could misclassify writing by people whose first language was not English. Ask the pupil to explain how they reached an answer, and consider drafts or other evidence under the applicable assessment rules. For qualifications covered by JCQ, its guidance addresses authenticating work and how permitted AI use should be acknowledged. Can You Tell It's AI? discusses the limits of detection in more detail.[6,7]

For a skill pupils need to be able to use independently, I would keep a part of the lesson and the assessment unaided. Suppose a pupil gets stuck on an equation. Let them try it first, ask for a hint, and then solve a fresh problem without the tool. The fresh problem tells the teacher something the first answer alone cannot.

Sources and notes

[1] Elizabeth Ligon Bjork and Robert A. Bjork (2011), Making Things Hard on Yourself, But in a Good Way: Creating Desirable Difficulties to Enhance Learning, in Psychology and the Real World, pp. 56–64.

[2] Evan F. Risko and Sam J. Gilbert (2016), Cognitive Offloading, Trends in Cognitive Sciences 20(9), 676–688.

[3] Hamsa Bastani and colleagues (2025), Generative AI without guardrails can harm learning: Evidence from high school mathematics, PNAS 122, e2422633122.

[4] Zara Contractor and Germán Reyes (2026), Experimental Evidence on the Learning Impact of Generative AI, IZA Discussion Paper 18792.

[5] Greg Kestin, Kelly Miller, Anna Klales, Timothy Milbourne and Gregorio Ponti (2025), AI tutoring outperforms in-class active learning, Scientific Reports 15, 17458.

[6] Weixin Liang and colleagues (2023), GPT detectors are biased against non-native English writers, Patterns 4(7), 100779.

[7] Joint Council for Qualifications, AI Use in Assessments: Protecting the Integrity of Qualifications.