# AI in education

22 June 2023

**Written by** Simon Peyton Jones

These notes were originally written at the invitation of the Royal Society's Education Committee, as a discussion starter for a conversation about AI in education at school (ages 6-18), especially in the light of recent progress with large language models (LLMs) such as ChatGPT.   This blog post is a lightly edited revision.

1. Don't conflate

When talking about "AI in education" It is really important to distinguish two very different things:

* (A) **Teaching *with*AI**: using AI as a piece of educational technology, to improve teaching and learning.
* (B) **Teaching *about*AI**: teaching students about AI itself, so that
  + they can make well-informed choices about how and when to use AI
  + they can make effective use of AI tools
  + they can get jobs in the rapidly-expanding AI field

Conflating (A) and (B) leads to endless confusion.  Let's not do it.

2. AI as ed-tech

2.1 The ed-tech opportunity: intelligent tutoring

It is just possible that AI may open up a major opportunity in teaching and learning.  There is [substantial evidence](https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/one-to-one-tuition) that 1-1 tutoring is highly effective in helping students to learn faster.  What if that tutor was an AI system, rather than a human being?  Until now, we have had only clunky, virtual learning environments, with myriads of hand-crafted pathways, but that are inevitably much less responsive to the specifics of the individual student than a human would be.  But ChatGPT has demonstrated a kind of step change in AI's ability to respond in a truly meaningful way to a prompt.

There are plenty of "personalised learning" systems out there, which adapt in some way to learner's progress, but large language models change the game.  It is now imaginable that an AI system really could make a truly decent attempt at understanding what problem a student was having, giving them hints and suggestions, identifying what misconceptions they hold, and enlisting a human being when things aren't going well.  It is even conceivable that an AI tutor could be *more*effective than a human one, in some ways: getting it wrong in front of a machine is less embarrassing than doing so in front of your maths teacher.

It is yet to be demonstrated that this would work in practice, but if it did it could have a transformative impact.

There is a useful summary of the state of the art in Intelligent Tutoring Systems in the Unesco report [AI and education: guidance for policymakers](https://unesdoc.unesco.org/ark:/48223/pf0000376709) (Section 3.1); it precedes the ChatGPT inflection point. By way of an example,  [Khanmingo](https://www.khanacademy.org/khan-labs)is the Khan Academy's ITS.

2.2 Challenges for assessment

Love it or hate it, education currently involves summative assessment.  ChatGPT has aced some quite challenging exams, which threatens remote forms of summative assessment..   My guess is that LLMs will lead to a continued emphasis on "traditional exams", with all their strengths and weaknesses, in which students are physically present.

Formative assessment is more of a problem.  If students cheat they harm only themselves. And yet they will cheat.  Moreover, in many schools, formative assessment bleeds over into summative, because it is used to monitor progress.  This is already a real problem for teachers, right now.  ChatGPT is writing English essays.   In a sense this is not new: my children often use search engines and then copy-paste stuff into their answers; and much homework is done in collaboration with, or by, parents or siblings.    But LLMs do seem to represent a step change.

One question people ask is: if you are setting formative assessments a robot can complete, shouldn't you change your assessments?  Well, no.  What we teach and how we teach will be *affected*by AI, but consider this: we still teach numeracy for years at primary school, even though a humble calculator can do the same tasks far faster and more reliably than any human.  There are sound educational reasons for doing this, ones to do with human cognitive development, and not to do with robots.   [Daisy Christodoulu works through this question](https://substack.nomoremarking.com/p/if-we-are-setting-assessments-that-a-robot-can-complete-what-does-that-say-about-our-assessments-cbc1871f502) far better than I could.

3. Teaching young people about AI

3.1 Avoid knee jerk

Back in 2010, "mobile" was all the rage. Yet we carefully avoided putting that word into the programme of study for computing, because what is hot at one moment may not be hot ten years later.   So we should not too readily toss our education system in the air, to respond to the AI boom.

That said, I think one can make a strong case that this particular shift will change fundamentals, and isn't just a passing phenomenon.  Here are two ways in which AI might have an enduring effect on the fundamentals of computing education:

* The term "computational thinking" became very popular when thinking of computing education.  But computing is the study of computation, information, and communication.   AI promotes a much more data-centric, rather than computation-centric way of thinking: perhaps "informational thinking" rather than just "computational thinking".  Perhaps that is a welcome rebalancing.
* Faced with the task "please get a computer to do this job", we have until lately always responded by writing a program.  But now we can respond by getting the computer to learn from data.  That's a *fundamentally*different way of approaching "get a computer to do this job", far, far more so that "program in Python" vs "program in C".

Still, I think we should be cautious about how rapidly or radically we seek to change what we teach, or how we teach it.  Education is complicated.  It is very easy to cause unintended consequences.

Moreover, even if we made it a Top Priority to produce more graduates with excellent AI skills, that does not mean that we should up-end the school curriculum for computing.  To understand AI you need a good grasp of the fundamentals of computer science, and that is precisely what the National Curriculum aspires to do.  We might discuss how to do it better; we might tweak it a bit (e.g. a bit about data, a bit less about algorithms; a stronger emphasis on statistics and probability in maths) but I think that up to age 16, tweaking should be enough.

Post 16, when students exercise greater choice and specialise more, one might envisage more substantial changes.  For example, an A level in Data Science could be an attractive marriage of maths, statistics, and computer science.

3.2  It's just a trillion floats

From the outside, AI looks like outright magic.   I know quite a bit about how it works, and I'm still totally astonished by recent progress.

But magic is dangerous.  If a young person has *literally no idea*how a computer system works, they are more prone to

* Trust it ("the computer said it so it must be true")
* Anthropomorphise it ("Does ChatGPT have feelings?")
* Use it in situations when it is an inappropriate tool.

I think these risks are mitigated, to some extent, once you realise that ChatGPT is no more than a bank of a trillion floating numbers, and lots of multiplies and adds.  How can a trillion floats have feelings?

So I think there is, arguably, merit in giving our young people an elementary understanding of how neural networks work, the sort of understanding you can get in a few hours of study (e.g. [this talk](https://www.youtube.com/watch?v=Q8PuC9nY15g)).  I stress *elementary*understanding.  I want to shine just enough daylight on the magic so that it becomes not magic but just more technology.

3.3 Educating critical judgement, avoiding FUD

One point that everyone has taken on board is that AI is rife with both opportunity and risk.

On the risk side we can think of

* **Bias**: if the system is trained on unrepresentative data it will cough up biased or unrepresentative answers.  (Much has been written about this.)
* **Confabulation/hallucination**: the system confidently makes up stuff that isn't true, effectively presenting fiction as fact.
* **Explainability**: "*you can't have your insurance because computer says no; sorry, I can't tell you why*".
* More existentially: **how might AI affect society**: jobs, even AI-takeover. [Geoff Hinton's recent observations](https://www.bbc.co.uk/news/world-us-canada-65452940) (and he is a quintessentially well-informed critic) are pretty strongly worded.

Everyone nods sagely and agrees that our young people should learn something about the risks and opportunities of AI.   But the consequences of AI are sufficiently profound that we really need to make sure it happens, and for all our young people, not just the geeks.  As Jefferson said: "a well-informed citizenry is the best defence against tyranny".

Moreover, it makes sense to balance the risks with the *opportunities*-- beyond the "how to cheat at your homework" opportunity that students see.  For example, using a LLM to search for information on a topic is probably much better than using existing web-search tools.  "Better" means: finding more relevant information, more quickly, better organised, and (importantly for young people) ad-free. Yes, the answer might be incomplete, biassed, or outright wrong.   But existing search engines turn up plenty of hits that are irrelevant, misleading, or even downright harmful.  It doesn't have to be perfect to be better; and in both cases you need to exercise critical judgement in making sense of the results.

Discussions about the opportunities and risks of technology already have a place in the existing [national curriculum for computing](https://www.gov.uk/government/publications/national-curriculum-in-england-computing-programmes-of-study) but, given the febrile hype that surrounds AI at the moment, there is a real danger that all we achieve is to create a fog of [fear, uncertainty, and doubt (FUD)](https://en.wikipedia.org/wiki/Fear,_uncertainty,_and_doubt), that *inhibits*well-informed debate rather than *enabling*it.

Schools and teachers may need help in framing and facilitating such discussions so that they are well-informed rather than merely alarmist.An analogy: it is reasonable to be concerned about what we eat.  But it's unhelpful simply to say "there are lots of dangerous chemicals around, you might get cancer" etc.  Much better to be more specific; to be evidence-based (if you smoke 10 cigarettes a day, your life expectancy is typically reduced by 3 years); and to give concrete and actionable suggestions for healthy eating.

4. Are LLMs an inflection point?

Tech has always affected education.

When calculators became cheap, maths educators worried that they were a threat to numeracy education; when search engines became ubiquitous, students could do homework by search/copy/paste;  online translation bots speak excellent French; when students can communicate rapidly and invisibly, cheating is easier.   And yet we still teach numeracy; we still teach French; we still set homework.  All of these bits of tech are opportunities too, and we have learned to treat them as such.

So are LLMs just more of the same?  Should we just say "calm down, soon it will all be normal and fine"?   It is possible to argue both points of view:

* **It's a step change.** LLMs are a real inflection point, a qualitative shift, not "more of the same". By passing some invisible waypoint of scale, LLMs can now do things that seem *qualitatively*different and remarkable, such as assembling a coherent sequence of points into an argument.  And progress is rapid.  Maybe there is a lot more to come.
* **It's just another increment.**  Yes it's amazing, but we are already close to a plateau. ChatGPT was trained on most of the internet -- there just isn't much more data to feed it.  It cost $100m to train; that's a lot, and spending (say) 100x as much seems implausible.  
    
  And, even aside from confabulation, it clearly lacks "understanding".   My favourite example of this is due to Stephen Wolfram [who asked](https://writings.stephenwolfram.com/2023/01/wolframalpha-as-the-way-to-bring-computational-knowledge-superpowers-to-chatgpt/) "*How many calories is in a cubic light year of ice cream?*".  To which ChatGPT replied in beautifully-idiomatic English: "*I'm sorry, but you can't have a cubic light year of anything, let alone ice cream*".  It doesn't "understand" that you can turn *any*length into a volume.

Personally I'm in the "inflection point" camp, for now anyway.  But I think it's a good debate to have, if only because it may lend perspective, helping to avoid getting carried away with the hype.

5. Reading list

In a hype-filled space, pointers to thoughtful and well-evidenced writing is valuable.  Here are some that I have gleaned so far.   Please send me others, or add them in a comment on this post.

* [The DfE guidance on generative AI in education](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1146540/Generative_artificial_intelligence_in_education_.pdf) (March 2023)
* Unesco: [AI and education: guidance for policymakers](https://unesdoc.unesco.org/ark:/48223/pf0000376709) (2021).
* [Stephen Wolfram on ChatGPT](https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/) (Feb 2023)
* [teachai.org](https://teachai.org/) is "committing to provide thought leadership to guide governments and educational leaders in aligning education with the needs of an increasingly AI-driven world and connecting the discussion of teaching *with* AI to teaching *about* AI and computer science".   This "with" vs "about" is my (A) and (B) above.
* [Sparks of Artificial General Intelligence: Early experiments with GPT-4](https://arxiv.org/abs/2303.12712).  Bubek et al, Apr 2023, ArXiv.  This paper is written by very well-informed people, but takes a "maximalist" view, with which many disagree.  To quote the abstract:  *"Given the breadth and depth of GPT-4's capabilities, we believe that it could reasonably be viewed as an early (yet still incomplete) version of an artificial general intelligence (AGI) system."*
* [The opportunities and risks of foundation models](https://arxiv.org/pdf/2108.07258.pdf).   Bommasani et al, Stanford, July 2022, ArXiv.  Long (200 pages) but measured and thoughtful.  *"This report provides a thorough account of the opportunities and risks of foundation models, ranging from their capabilities and technical principles to their applications and societal impact."*
* [Generating harms: generative AI's impact and paths forward](https://epic.org/documents/generating-harms-generative-ais-impact-paths-forward/), Electronic Privacy Information Centre, May 2023.  This report gives a detailed breakdown of the possible harms that may arise from generative AI, with an extensive bibliography.