

An introduction to the role of Artificial Intelligence in classrooms and schools



the
education
hub

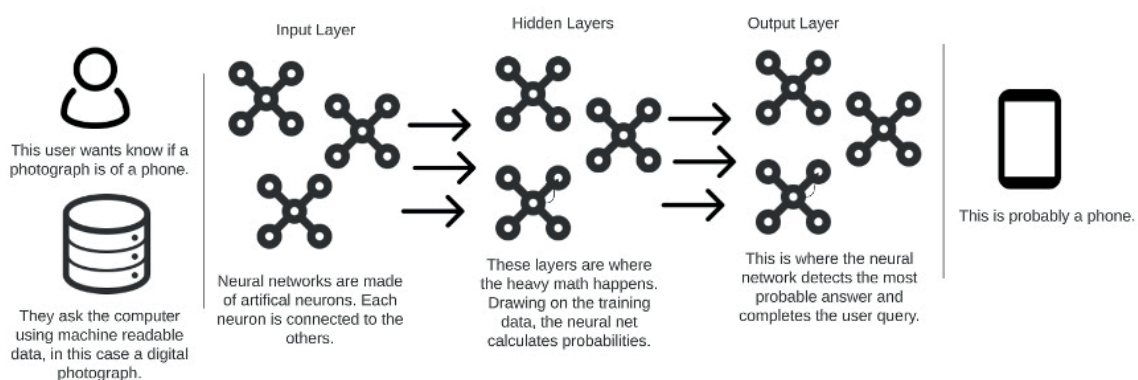
School resources

In late 2022 and early 2023, a series of articles and features broke across popular media. Artificial Intelligence (AI) had arrived, and everything apparently had changed. This research review will describe what is meant by 'artificial intelligence', what it does, how it works, and what the evidence says about the implications for education.

What is Artificial Intelligence?

Artificial intelligence or AI is a term used in computer science to refer to processing tasks that require considerably more advanced data synthesis and analysis than those used in ordinary computing applications. AI comes in two primary forms, algorithmic and neural networks. Until very recently, consumer facing applications of artificial intelligence have tended to be algorithmic, or rules-based. An **algorithm** (words in bold are in the glossary at the end of the review) is basically a series of 'if/then' rules, structured for a machine to follow. Simple algorithms in everyday life include things like sewing patterns or recipes. For computers, these rules are sometimes explicitly written by programmers and sometimes the rules are 'learned' by the computer from large data sets. Complex algorithmic systems usually mix these explicit (written) and inferred ('learned') rules.

However, **neural network** AI systems work differently. These are the kind that have resulted in this most recent surge of new products, such as **Chat-GPT** for text and **DALL-E** for images. These systems are modelled on how biological brains work. Rather than being given a set of explicit rules, these systems are 'taught' by looking at the patterns in relationships between things, with classification guidance provided by programmers. These systems require massive amounts of training data, on the scale of the whole internet. Posting on social media or using a Captcha contributes training data for an AI product of this type.



Familiar voice assistants, including Siri, Alexa, and Google Assistant, tend to be algorithmic systems. These systems are more rigid and their responses are usually predictable and repetitive. They often fail when they encounter a query they have not been trained to understand. While they are programmed to recognise and imitate natural language, their responses are scripted so people quickly discover the limitations of these systems when using them. These systems are unlikely to be mistaken for anything

other than handy (and sometimes frustrating) call-and-response machines. Many of the pop-up chat-bots on the websites of banks and retail shops are of this type too.

Academics and scientists have been experimenting for a while now with a different AI that learns about the connections between 'things'. These are neural net AIs. These systems are not provided with rules for making decisions, but rather 'learn' rules for how to understand connections and basic context from within a set of data. What these machines then understand as the relationship between 'things' determines how they produce an output.

Data, data, data

Neural net AIs require enormous data sets. These data sets include significant proportions of the entire internet, humanity's written words, both conversational and formal, and all the images imaginable. In the latter half of 2022 the 'third generation' of many of these AI experiments were made available online for people to interact with.


The two primary kinds of these third generation AIs that are relevant to this article are both **LLMs** (Large Language Models), where the 'language' involved is text or images. The outputs of these, again either images or text, are called [Generative AI](#). LLMs can generate conversational language and take in requests from people in natural language at a level never seen before. But more importantly, they can generate content at a quality level that has profoundly surprised many users.

Immediately this raised concerns amongst teachers and educators. First, AI-generated images from the likes of **DALL-E** and **Stable Diffusion** shocked people with the ability to create images at a level of quality and accuracy they had not before experienced from short descriptive prompts.




Created by DALL-E in response to the prompt, 'Can you show me a kiwi bird?' Generated on 2 March 2023.

But following along from these, attention quickly shifted to an AI product called **Chat-GPT**. Chat-GPT can generate text in a huge variety of styles, in different genres, fiction and nonfiction. It can cite, make arguments, connect facts, construct introductions and conclusions, using a variety of tones, from journalistic and conversational to formal and academic. However, first appearances were misleading, as it turned out that the nuance of what Chat-GPT produced lacked rigour. The devil, it turned out, was quite literally in the details. This will be covered in more detail below.



Can you write a lesson plan to teach about verb tenses to a group of 13 year olds.



Sure! Here's a lesson plan to teach verb tenses to a group of 13-year-olds: 👍 🗨

Topic: Verb tenses

Objective: Students will be able to identify and use simple present, simple past, and simple future verb tenses in context.

Materials:

- Whiteboard and markers
- Worksheets with verb tense exercises
- Examples of sentences using simple present, simple past, and simple future tenses
- A short story that uses all three verb tenses

Warm-up (5 minutes): Begin the lesson by asking students to brainstorm verbs they know. Write their suggestions on the whiteboard. Then, ask them to identify what the different verb tenses are and give examples. Write these examples on the board as well.

Introduction (10 minutes): Define verb tenses as the different forms of verbs that show when an action happened or will happen. Explain that there are three basic tenses: simple present, simple past, and simple future.

Example from Chat-GPT3, generated on 2 March 2023.

In real life, neutral net AIs also include algorithmic aspects. Developers and researchers have seen the limitations in the current generation of AI and are placing controls on the LLMs through algorithmic moderator bots intended to make these tools safer. This does not include preventing what they are used for, which will be covered more below. In current practice, this can be seen when an AI comes back with a response that is formulaic, simplistic, lacking a conversational nature or vagueness, and resembles something out of a smart-speaker. These new releases of AI are certainly impressive, but people familiar with them know that they most definitely have limits.

Does this herald the death of the essay?

Whenever a shift in technology occurs there is always considerable hype about the transformative nature of it, although most of this tends to be overblown. There is no evidence that the recent developments in AI will be any different to previous waves of hype. Also, it is important to realise that none of this is especially new. It works like auto-fill search fields and auto-correct. The scale is certainly different, and its abilities have expanded considerably beyond what has been experienced up to now, but importantly this is not a change in kind. Students are already using **Grammarly** and similar tools, and **Chat-GPT** is not far removed from this.

This is not to say that there will not be changes as a result. As mentioned above, many of these AI developments are in their third 'generation' or iteration. Some are now even into their fourth. More and more generations of these are expected to be developed, especially given that this technology is starting to move out of the research and academic arena into the private sector. The implication of

these iterations is that the flaws that generative AI currently display will become less and less apparent. What we know now to look for in determining whether or not something is AI-generated will no longer be reliable indicators as the systems improve.

Still, what is generated by these AI systems is at a basic level convincing. The outputs sometimes lack depth and expertise, and it is often this lack that indicates that the content may not be quite right. In other words, they have reached being 'good enough'. But what they are 'good enough' for is the fundamental question that teachers and educators need to be asking.

Questions of quality

A primary concern is that AI tools can be wrong, sometimes considerably so. The tone and format of an AI may read like it is sure of the facts and arguments it is making, but the facts and arguments can be quite incorrect, and without any of the signals that humans who are uncertain or confabulating send. This occurs because it is trained on data, and the tool is dependent on that data being correct. If something incorrect is in the data with sufficient regularity, the connections that it picks up will be repeated. Further, many questions asked of these tools are filled with complexity: relationships between relationships. The material used will tend to be relevant, but can be lacking in robustness, or even just explicitly false. Note that this is not what is often called an **AI hallucination**. An AI hallucination happens when the AI diverges from the data or is unfaithful about it. This is another way that AIs can [fail](#).

Further, what is typed into a prompt can be conflicting for the AI. There are many documented cases of a person asking an AI to generate an essay with references. The AI will do this, but it sometimes creates an entirely invented [citation](#) which is convincing, but fake. A student using generative AI to write an essay could easily be fooled, not realising that they may be compromising their academic integrity. Even if the reference actually exists, what the referenced author states may not be what the AI generates, or what the author has stated may not be able to be used to support [the claims the AI has generated](#). These are crucial questions to remember, as they form part of how a response to this technology could be framed.

Moreover, and importantly to this point, the AI doesn't know what is true and what is false. All AI tools are limited to what is in the databases of information they have been trained on. The information they are providing means nothing to them. Meaning doesn't matter, and true and false are beyond them. Similarly, contextual particularities and less common nuances not only do not matter to them, but they are not aware of them. This includes the connections and patterns beyond the basic ones they develop, the ethical or moral concerns about what the information conveys, and questions of import regarding the information.

It has also been noted that the ability of these generative AIs to write essays convincingly does vary depending on the purpose of the essay. On those topics that require higher levels of critical thought, interpretation, and applications of theory for analysis, particularly where context matters, [AI tools do not perform as well](#) as topics where these are less central. Overall, though, the essays produced do tend to be 'good enough'. They might result in a C or a B, but lack that excellence that would result in an A. However, a C or B is still a good passing grade, and so, despite the above concerns regarding accuracy, this is something that teachers need to reckon with and develop good strategies and approaches.

What this means for teachers and schools

First, teachers do not need to worry about being replaced with an AI teacher. That scenario is a very long way off¹. Some institutions have taken the approach of [banning student use](#) of these tools. These institutions consider it in the same arena as plagiarism and the passing off of another's work as a student's own. TurnItIn and similar companies have [announced](#) that they will be rolling out feature

additions to their existing software to detect AI-generated content. How precisely this will be done remains unclear.

While banning the use of generative AI is an option, this could quickly become a moving target as AI keeps improving. Constant policing to detect AI use will be needed in order to keep pace with developments in AI over time, demanding staff resourcing and financial investment. Instead, an approach that recognises that the use of generative AI tools by students is happening², and [plans](#) for this use, offers a more practical option.

What do we know?

When considering how to move forward, age matters a lot. These tools are not the same as algorithmic driven chatbots and there is strong evidence that they can generate [toxic outputs](#), even with 'mod bots'. Chat-GPT appears to be less likely to produce such outputs than the AI version of **Microsoft's Bing**, which is still only available to certain users, although Chat-GPT has also had toxicity in its [outputs](#). These models are not controlled in the same way that algorithmic chat-bots are, and they present a different risk profile. While there is a small and growing evidence base around using AI tools in early childhood education³, this evidence base pertains to algorithmic AI tools, not generative AI⁴.

The evidence-base around generative AI tools in education is extremely limited. They are simply so new and are developing so quickly that there has not been enough time for the scholarship around them to develop. The few studies that exist are small, quite narrow, and tend to use specialised deployments that sidestep toxicity issues. They also, of necessity, cover short term projects rather than offering a longer term look at the value of these tools for students and teachers. That said, there is an [evidence-base](#) around the use of algorithmic chat-bots⁵. It is not yet known if what we have learned about those chat-bots applies to a product like Chat-GPT.

Supporting students to learn about artificial intelligence is going to require coordinated approaches across the curriculum and year levels. The concept of **AI literacy** has led to the development of a flurry of scoping work around what this means and how to deliver it within curricula⁶. Singapore, South Korea, Australia, and the EU have begun their work to incorporate teaching about these tools into curricula⁷. [Canadian schools](#) are also developing coordinated approaches. New Zealand will need to do the same.

Practical approaches

The first thing to do is to talk to students about these new AI products, their limitations, what they can produce and what they cannot. Students are going to find out, and an approach that relies on them not knowing about it will result in them using it badly. Of course, do so at an age-appropriate level, but let them know about it and what this could mean for them. Demonstrate how they work in class, and discuss limitations and potential uses. A [2021 study](#) which examined the process of developing a curriculum with a learning trajectory for these sorts of AI might provide a sense of how to approach this conversation in a systematic way and help to identify learning objectives. It is also worth being careful around disparate impacts. One study showed that using these tools in a painting class left girls less engaged than boys⁸.

If teachers choose to tell students they are allowed to use generative AI tools for their assignments, it is important to tell them how to use it and talk about their limitations. Some of them will be using it anyway, so best to be direct about expectations. There are a few studies that offer guidance on how these tools can be used in the classroom and their utility as a learning tool. For example, [researchers explored](#) the question of whether a special implementation of Chat-GPT could be used to encourage students to develop and ask more curious questions. The study involved 9- and 10-year-olds, and used a specialised version of Chat-GPT with strong protections to avoid toxic outputs (it is not currently available for

public use). The study found that this version of Chat-GPT was useful helping the students to ask better questions. A recent [position paper](#) lays out a range of other uses of this technology for both students and teachers. Again, it is early days, but the range of possibilities is wide.

Rethinking assignments

Given that AI performance varies depending on the subject being taught, it might be good to explore this further. As mentioned, questions that encourage interpretation, critical analysis, contextualisation, and comparative analysis are less vulnerable to AI misuse. When showing students how generative AI tools work in class, try different ways of framing a prompt to show them how to investigate something. Get them thinking about a good prompt and what might go into it. Relate the different prompts to the different parts of the research process, or use the prompts to investigate the different parts of the research question the students need to answer. Particularly to that point, it is important to rethink how research questions are framed. For older students, emphasise research and analysis skills over 'find me the facts around x' essays. If essay framings and use-cases are having to shift, do so in such a way that encourages focusing more on critical thought and analysis.

The benefit of these approaches to incorporating AI into teaching is that they help future-proof for advances and improvements. While it is tempting to ban any use of this technology, not many institutions will have the ability to enforce these bans. As advances in the abilities of this technology almost inevitably occur, designing teaching around what it can provide rather than attempting to avoid it will not only allow teachers and schools to be more future-proofed and flexible, but it will ensure students can better manage their use more skilfully outside and after class.

Are there ethical concerns?

As with any new technology, academics and researchers have ethical concerns. One of the largest such concerns is around the [legal status of the data](#) that is used to train these AI. There is currently a lawsuit pending from Getty Images against Stable Diffusion, one of the biggest image-generating AI, and more lawsuits from other media companies are expected. However, while large media companies have the resources to bring copyright lawsuits, many small creators and artists do not. The crux of the position here is that their work is being used, without consent, knowledge, or recompense. Further, styles of art that may be specific to an artist can be (re)produced by this technology. While the focus here has been on image generating technology, it is no different for LLMs that focus on the written word. Undoubtedly this is something that will be worked out in the courts over time, but there is no guarantee that smaller creators will be respected. Any use of the technology has to be with this in mind.

Bias in the outputs of these tools is also a worry. There has been considerable research done on the ways in which [bias is replicated and made worse](#) by AI technologies. There are two common ways in which AIs can be biased: in design choices (either implicitly via ignorance or assumption, or explicitly via prejudice) or via bias in the sets of data the AI is trained on. The former is clearly evidenced in decisions made by early facial recognition technologies that were not given information on darker skinned faces. Neural net AI tools are particularly susceptible to training data bias. These tools are trained on enormous sets of data that are pulled from online sources and some of that data comes from the less savoury corners of the internet. If bigoted and discriminatory training data are common enough within an AI tool's data, the AI will replicate that in its outputs.

Further, there is the potential for misinformation to be spread by this technology. AI tools can be used by bad actors to generate large amounts of hateful or even just incorrect text. This text can then be shared widely, at a much lower cost than having human writers generate misinformation. Also, given

the credence the content produced by AI has, it lends itself to having a degree of authority. This is problematic given their inability to determine meaning or whether something is correct.

The credibility the technology has as an information source is related to another ethical concern, that of anthropomorphism. Humans have a tendency to imbue animals and things in their environment with a degree of 'humanness'. This tendency has historically been the domain of culture and religion, and it has been a rich source of inspiration about the world and how we understand our moral relationships. Many AI tools have been designed to appear human in a way that encourages this tendency. While some AI tools, like those offered by Character AI, do so clearly within the realm of fiction, most do not. This means that it is easy for people to see the AI as [self-aware or having agency](#). This is deceptive and can lead to serious harms, including radicalisation to violent extremism, that are [still being explored](#) in the scholarly literature. There is also a potential risk in humans transferring their poor treatment of AIs to human beings or animals. If we treat an AI poorly and we see that AI as being 'human' to an extent, then negative patterns of behaviour towards the AI can be replicated in our dealings with people⁹.

Finally, we need to consider the impact on climate change mitigation efforts. AI requires processing power that is considerably larger (due to their higher and different levels of processing) than current data processing centres, which themselves already have an [impact on our power grids](#). Widespread adoption and ubiquity of AI will result in larger amounts of resources required to run them, and it is important to be aware of how this will affect our electricity use.

Note: this technology is rapidly advancing, so this research review must be located in the context of its time and is valid as of early 2023.

Further reading and resources

[ChatGPT is everywhere. Here's where it came from | MIT Technology Review](#)

A solid history of the development of Chat-GPT and how we got to where we are.

<https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web>

A thoughtful and sceptical article about Chat-GPT and its potential.

[\[2301.04655\] ChatGPT is not all you need. A State of the Art Review of large Generative AI models](#)

This article offers a taxonomy of the primary models used currently for generative AI tools.

https://edarxiv.org/5er8f?trk=organization_guest_main-feed-card-text

A position paper on the potential uses of generative AI in education.

[What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education | Smart Learning Environments | Full Text](#)

This article offers a summation of the current uses of chat-bots in education and offers a research agenda.

[On the Dangers of Stochastic Parrots | Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency](#)

A comprehensive analysis of the ethical concerns and risks of these tools.

Glossary and product directory

AI hallucination

[https://en.wikipedia.org/wiki/Hallucination_\(artificial_intelligence\)](https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence))

AI Literacy

<https://www.sciencedirect.com/science/article/pii/S2666920X21000357>

<https://www.bcu.ac.uk/education-and-social-work/research/news-and-events/cspace-conference-2021/blog/ai-literacy-the-role-primary-education>

Anthropomorphism

<https://en.wikipedia.org/wiki/Anthropomorphism>

Artificial Intelligence

https://en.wikipedia.org/wiki/Artificial_intelligence

<https://www.britannica.com/technology/artificial-intelligence>

Algorithm

<https://en.wikipedia.org/wiki/Algorithm>

<https://www.britannica.com/science/algorithm>

Character AI

<https://beta.character.ai>

Chat-GPT

<https://en.wikipedia.org/wiki/ChatGPT>

DALL-E

<https://en.wikipedia.org/wiki/DALL-E>

Generative AI

<https://www.techtarget.com/searchenterpriseai/definition/generative-AI>

<https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai>

Getty Images

https://en.wikipedia.org/wiki/Getty_Images

Grammarly

<https://en.wikipedia.org/wiki/Grammarly>

Large Language Model

https://en.wikipedia.org/wiki/Wikipedia:Large_language_models

<https://www.forbes.com/sites/robtoews/2023/02/07/the-next-generation-of-large-language-models/>

Microsoft's Bing

https://en.wikipedia.org/wiki/Microsoft_Bing

Neural Net

https://en.wikipedia.org/wiki/Artificial_neural_network

<https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>

Stable Diffusion

https://en.wikipedia.org/wiki/Stable_Diffusion

TurnItIn

<https://en.wikipedia.org/wiki/Turnitin>

Endnotes

- 1 Tack, A., & Piech, C. (2022). The AI teacher test: Measuring the pedagogical ability of Blender and GPT-3 in educational dialogues.
- 2 Hwang, G., & Chang, C. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, DOI: 10.1080/10494820.2021.1952615
- 3 Yang, W. (2022). Artificial Intelligence education for young children: Why, what, and how in curriculum design and implementation. *Computers and Education*, 3. DOI: <https://www.sciencedirect.com/science/article/pii/S2666920X22000169>
- 4 Su, J., & Yang, W. (2022). Artificial intelligence in early childhood education: A scoping review. *Computers and Education*, 3. DOI: <https://www.sciencedirect.com/science/article/pii/S2666920X22000042>
- 5 Su, J., & Ng, D. T. K. (2023). Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. *Computers and Education*. DOI: <https://www.sciencedirect.com/science/article/pii/S2666920X23000036>
- 6 Su & Ng, 2023.
- 7 Steinbauer, G., Kandlhofer, M., Chklovski, T., et al. (2021). A differentiated discussion about AI Education K-12. *Künstl Intell*, 35, 131–137. <https://doi.org/10.1007/s13218-021-00724-8>
- 8 Sun, J., Gu, C., Chen, J., Wei, W., Yang, C., Jiang, Q. (2022). A study of the effects of interactive AI image processing functions on children's painting education. In: Gao, Q., Zhou, J. (eds), *Human Aspects of IT for the Aged Population. Design, Interaction and Technology Acceptance. HCII 2022. Lecture Notes in Computer Science*, vol 13330. Springer, Cham. https://doi.org/10.1007/978-3-031-05581-2_8
- 9 Coghlan, S., Vetere, F., Waycott, J. et al. (2019). Could social robots make us kinder or crueller to humans and animals? *International Journal of Social Robotics*, 11, 741–751. <https://doi.org/10.1007/s12369-019-00583-2>

PREPARED FOR THE EDUCATION HUB BY



Mandy Henk

Mandy is a librarian, writer, and advocate for a healthy and vibrant Commons. She was named a 2012 Library Journal Mover and Shaker for her work as a founding member of the People's Library at Occupy Wall Street, and is the author of *Economy, Ecology, Equity: The Path to a Carbon Neutral Library* (ALA Editions 2014). Mandy currently serves as Chief Executive for Tohatoha, a not-for-profit organisation in Aotearoa New Zealand, that works within communities to help build solutions for an equitable and just internet, while amplifying the voices of those harmed by digital technologies.



Dr. Sarah Hendrica Bickerton

Sarah joined the Tohatoha team following nearly three years as a postdoctoral researcher and lecturer in Public Policy at the University of Auckland, and brings her deep academic background as a sociologist with expertise in technology, policy, and gender. Her PhD research was on political participation behaviours amongst women Twitter users.