A Guide to Generative Artificial Intelligence in Education

Authored by John Hackett on behalf of the NEN
with additional input from the Technical Strategy Group

Introduction

It seems that hardly a day goes by when Artificial Intelligence (AI) or Machine Learning are not somewhere in the news. At the time of writing (May 2023), most of the discussions are focused on what are known as Generative AI models. Two, in particular, have hit the headlines: *ChatGPT* from OpenAI/Microsoft and *Bard* from Google.

The term Generative AI (GAI) refers to models which generate content rather than provide expertise in a specific field like, for example, spotting cancer cells in X-rays. This is more correctly referred to as Machine Learning, a subset of AI.

In the case of both ChatCPT and Bard the output is text: a prompt is used to start the "chat", and the GAI generates a textual response. But there are other GAIs that generate audio (e.g. https://www.aiva.ai/) and images (e.g. https://openai.com/research/dall-e) in response to a text prompt or selection of parameters.

Why has AI suddenly become such a hot topic? And why should schools be interested and/or concerned? In this paper, we offer a broad overview of AI and the potential challenges for the education sector. We will consider both how AI has been used over the past few years and the impact that the rise of GAI may have on education and the wider society.

What is AI?

AI is a very general term which covers a wide variety of models or algorithms that try to mimic in some way how humans think. For the purposes of this paper, AI can be divided into two competing approaches. One is based on logic and reasoning. Generally, these are known as **Expert Systems**. In these AIs, a collection of rules are created that attempt to encapsulate a specific area of expertise, e.g. diagnosing a disease from a list of symptoms. In this example, the system will step the user through a range of questions using the expert system's rules to narrow down the options, ultimately giving both a diagnosis and a confidence level. The important point here is that the output is based on pure logic (stepping through the rules based on the answers to specific questions), and the steps can be shown and validated.

Expert Systems have had great success in specific areas, but for *each expert system*, a new rule set must be created by talking to human experts and trying to capture the expertise they have acquired through years of experience into a set of formal rules which can be programmed into a computer. This is both complex and time-consuming.

The second approach does away with the purely logical, rules-based method and instead takes its inspiration from the way the brain works. In the late 1960/70s, AIs started to be built using **neural networks**. In simple terms, a neural network is a self-modifying algorithm that "learns" by being "trained" on a set of data. The training data contains examples of both what is and is not the target. For example, if we want an AI to recognise images of dogs, then the training data will consist of images of dogs and not-dogs: the training consists of feeding back to the system whether it identified the image correctly. This feedback is critical in the training stage: it is this that "teaches" the neural network how to make the distinction that, ultimately, allows it to distinguish *dog* from *not-dog* on a new, previously unseen image.

This is **Machine Learning** (ML).

The key advantage that ML has over an expert system is that once you have a generic ML system, it can be trained on any appropriate data. While distinguishing between dog and not-dog is trivial, doing the same with skin cancer versus not-skin cancer is far from trivial. Once trained on a wide range of images of skin blemishes, the ML system can be used to help dermatologists decide whether a new image of a blemish is cancer or not. This example is used in a TED talk by Sebastian Thrun where he discusses what AI is and isn't.

Most ML (i.e. neural network based AIs) systems are trained to supply an answer in a specific domain (like the dermatological example above). However, the AIs making all the headlines now (May '23) - known as **Generative AI** (GAI) - are based on Large Language Models (LLM). An LLM is trained on vast amounts of data - typically scraped from the internet - and uses statistical methods to capture patterns, context, and the semantics of language to generate their naturalistic responses. It is very important to understand that these

are statistical methods and do not imply any understanding on the part of the AI.

It is, however, interesting to compare this to the famous "Chinese Room" argument in the 1980 paper by the logician and philosopher John Searle when arguing whether a computer can ever have consciousness or understanding.

The two most well-known examples of LLM based AIs are ChatGPT (from the Microsoft supported OpenAI) and Bard from Google. While ChatGPT and Bard are from different companies and have methodological differences, the underlying architecture and training methods are both based on the concept of the LLM.

GAI systems are not limited to text: there are GAI systems online now that can handle images (for example, the "Pufferjacketed Pope" image that went viral) and music (https://www.aiva.ai/). In these cases, the LLM is used to take the textual prompt to "understand" the request: for example, "create an image with the pope in a white puffer jacket" or "write a piece of sad, slow music with strings and piano". The output of this process is then used to generate an image or piece of music that matches the prompt.

The Positives

Having outlined what AI is and the three main types (Expert Systems, ML and GAI), we can begin to consider the potentially positive uses. In some cases, of course, what is a positive for one person may be a negative for another. There is no doubt that some (many?) jobs will be displaced as AIs become even more effective than the current generation.

AI is already being used around us: often in the background of everyday tasks. For example, the ranking algorithms in Facebook and Twitter, suggestions for which movie to watch on Netflix, which ads appear when browsing the internet, facial recognition systems, etc. These all use AI in one form or another to make decisions that affect our lives for good or ill.

It is also clear that AI has the potential to increase productivity in many areas of the economy. For example, according to Emad Mostaque (CEO of Stability AI) on Sunday with Laura Kuenssberg (14/5/23, 21'-29'), half of the code written now is AI generated, greatly increasing the productivity of programmers. Finding the optimal schedule for the transportation of goods is a complex problem, and logistics companies are already using AI to design their schedules and lower costs. In the same interview, he is also candid about the potential for job losses as the economy absorbs the imbalances that AI will create.

In the Law, AI is being used to speed up legal research with its ability to scan and analyse vast quantities of legal documentation and case law. Contracts can be analysed to assist lawyers in reviewing them to highlight potential issues and ensure their compliance with relevant laws.

In health, AI is being used to analyse images (X-rays, CT scans, etc.), to analyse electronic health records looking for patterns and trends, and in the analysis of clinical trial data, etc.

Even in the Arts, the use of AI is already being felt. In the US, the current dispute with the Writers Guild of America (WGA) is partly about the potential for using AI to generate TV scripts and reduce the fees to human scriptwriters. This dispute is an interesting case. The WGA is not after an outright ban on the use of AI but to have some regulations or "guardrails" on its use. They see that systems like ChatGPT can be useful tools for writing, but they are not "creative" as they are trained on pre-existing material, and are pushing back against the big media companies that want to *gigify* their work even more than it is already.

If we think about education in particular, there is a very interesting TED talk by Sal Khan of the Khan Academy where he outlines how they have incorporated AI into their systems. He explains how their AI works in two modes depending on whether the user is logged in as a Student or a Teacher. The AI does not give answers to students but prompts them by asking questions about their reasoning and allows the student to ask questions - much as they would to a human teacher. When a teacher is logged in, they are able to ask for the answer, and the AI can also offer suggestions on lesson plans and how to engage the student with the topic.

While this is limited to the Khan Academy users at the moment, it is likely that other online learning platforms are, or will be, working along similar lines and incorporating LLM-based AIs into their offerings.

These developments in AI may finally herald the age of true individualised learning, which has been promised for so long.

Even before this the current tools which have suddenly become publicly available, in particular **ChatGPT** and **Bard**, can be used in interesting ways within school. At its simplest, these GAI tools can be used in basic research: rather than using the formal search-term approach we have used up to now, questions can be asked in a more naturalistic way. I have used ChatGBT here to suggest some use cases! So here are a couple of its suggestions:

- *Suggest changes to written text, suggest ways to improve sentence structure or grammar, etc. See this blog post from Grammarly - an application which does a similar job - explaining how they use AI to suggest changes to written text.*
- *Students learning a new language can practice their skills by conversing with ChatGPT. They can engage in conversations, ask for translations, practice grammar and vocabulary, and receive feedback on their language usage.*
- *ChatGPT can serve as an additional participant in classroom discussions. Students can engage in debates, ask for different perspectives, and receive input on their arguments, fostering critical thinking and enhancing classroom engagement.*

These few examples demonstrate the huge range of possibilities for the creative use of the existing GAI systems for teaching and learning.

The game-changing advance from ML to GAI is the G for Generative. Machine Learning has been creeping into software for many years now and performs remarkably well (e.g. replacing a dull, grey sky with a beautiful blue one with white clouds or removing unwanted elements from a photograph). But the ability to create new images, to converse with the user, to write stories: that is new.

Those are the positives. But, as with most advances in science and technology, one must be on the lookout for the negatives and understand how to counter them. As the old saying goes: fore-warned is fore-armed. It is to these potential negatives that we now turn.

The Risks

As we have seen above, AI (both ML and GAI) provides teachers with powerful tools that can be used to enhance teaching and learning. However, it is important that educators are aware of the potential downsides of these new technologies. Just as teachers use the Internet as a tool but also need to understand and teach its dangers (pornography, gambling, grooming, etc) and put in mitigations (web-filtering, virus scanning, firewalls, etc.), so must they be prepared to teach about the dangers and limitations of AI.

Risks come in a variety of guises. There are risks inherent to the systems themselves and ethical risks. Then there are risks to society as a whole, which can be divided into the economic risks (worker displacement) and of criminals using AI.

Each of these risks has the potential to have a significant impact on society as a whole and will, inevitably, intrude into the education system.

Systemic Risks

By systemic risks, we mean those that are inherent in the system itself. There are two areas where such risks can be embedded: the software algorithms and the training data.

One well-known saying in computing is GIGO (Garbage In, Garbage Out). This can be applied to AI by one letter change - BIBO (Bias In, Bias Out). If the data used to train an AI is biased, then the resulting system will output biased results. If the training data is effectively the whole Internet (as is the case with the newest GAI models), then the AI will absorb whatever biases are out there.

A well-documented example of racial bias is the use of AI to assess the risk of a defendant reoffending. Mainly used in the US justice system, an investigation by Propublica showed that black defendants were more likely to be given a high risk of reoffending than white ones. Similar flaws have been found with facial recognition technology. The documentary Coded Bias (2020) is an investigation of the bias of ML facial recognition algorithms.

Both the examples above are ML systems - created to "solve" a problem - and the bias is either in the training data or the algorithms used to create the solution. Generative AI also displays bias but in potentially more subtle ways: is the image or text generated from the prompt based on biased training data and, therefore, itself biased? And how can you check? For example, a prompt of "surgeon at the operating table" to Stable Diffusion's demo system results in four images of white males in scrubs. Whereas "female tv presenter" gives four images of white, slim, mainly blond, 30-40-year-olds. Both are examples of stereotyping. Is this just reflective of the Internet's own bias?

Ethical Risks

With ML systems, there are two categories of ethical risk. Firstly, how has the training data been curated? Has due care been taken with the accuracy of the data, and have biases in the data been accounted for and mitigated? It is sometimes assumed that algorithms are inherently neutral and rely purely on reason, but this is not the case: programmers have biases (even if they are unaware of them) which may leak into the algorithms they create.

Secondly, there are ethical questions around how ML systems are used. Facial recognition by the police in public places, for example, has proved to be controversial.

The training data for ML systems tend to be discreet, curated datasets designed to train the system to undertake a specific task: cancer scans for diagnoses, war-zone satellite images for intelligence analysis, images of faces for recognition systems, for example. By contrast GAI systems (and the LLMs they are based on) are, in general, trained on vast quantities of data scraped from the Internet that will not have been checked for bias.

An ethical area that is of concern for GAI systems is the use of copyright material without the copyright holder's permission. Just because an image, for example, is available to view on a website does not mean that it can be used in other contexts without asking permission: there is a range of copyright licences

that restrict how freely a work can be used. With a Creative Commons licence the copyright holder may allow a work to be used freely for non-profits, for example, but require a fee for commercial use. The dispute between Getty Images and Stable Diffusion is one example of a copyright dispute in this area. The New Scientist magazine ran an article in their 13/05/23 issue about legal challenges to OpenAI's use of web scraping copyright material to create the training dataset without permission or fee.

The use of copyright material for training leads directly to the ethics of using a GAI to imitate a particular style by, for example, directing the GAI to generate a musical track "in the style of" a pop star. Or an image in the style of an artist? Is there an ethical difference between an art student looking at art and then making a piece in the style of, say, Picasso and an AI doing the same? This is likely to end up in the courts!

Societal Risks

Societal risks fall into two categories, the second of which will be dealt with in the next section ("Bad Actors"). The first category of risk arises from the legitimate use of AI and the impact that it will have on society. Automation and its close cousin Robotics have long used AI in their design. Recent advances in AI will inevitably lead to more efficient and versatile robots undertaking a wider range of tasks, from mainly mechanical jobs like fruit picking and car assembly to jobs too dangerous or impossible for a human worker, such as nuclear power station inspection and bomb disposal, to "soft" jobs like nursing and dementia care. See this video from Boston Dynamics to see the versatility of their *Spot* robot (or this one for a more humorous take!). Whether this is a good thing or not is open to debate: but what is certain is that AI will increase the spread of robots into new areas that will have an impact on society and the jobs available to the general population.

Bad-Actors

Next, we come to the "Bad-Actors" risks which have, perhaps understandably, gathered the most attention with the arrival into the public sphere of GAI in the form of ChatGPT and Bard.

What has been most striking about these GAIs is their ability to generate content that is "good enough" to convince the reader/viewer/listener to think that it was created by a human being. If we consider text specifically (but understanding that the same will apply to music and images), there is certainly the potential for a GAI to be prompted to create content about a controversial issue and automatically post it on social media. This may be done for overtly political reasons (e.g. posts about abortion, the US Jan 6th events at the Capitol, or Trump's claims that the election was stolen) or as "clickbait" for purely financial gain.

It is certain that in time (if they are not already), criminals will be using these tools to craft more persuasive text or voice messages to run scams.

Deep-fakes are another area ripe for political and financial exploitation. While "Deepfake Tom Cruise" may just be a bit of fun, the potential for malicious use of deep-fake technology is obvious: from scamming adverts, influencing elections, and revenge porn.

LLM "Hallucinations"

Large Language Models (LLMs) are the basis for GAIs. It is important to reiterate here that these models are, at heart, purely statistical. The LLM is trained on a vast quantity of data that is analysed by the complex algorithms that form the LLM itself. What the LLM is "learning" is statistical relationships between bits of text. The key point is that there is no understanding of either the input prompt or the output.

Nowhere is this better demonstrated than in the so-called LLM "Hallucination" effect. Or, more prosaically, an LLM sometimes *just makes things up*! Because there is no inherent understanding or "real-life" context to what the LLM is doing (it is essentially just manipulating symbols based on statistical rules), sometimes it will create a sentence that is grammatically perfectly acceptable but is clearly false. For example, Duke University has posted an explicit warning to its students that using ChatGBT as a research assistant has serious limitations: including fabricating convincing looking citations.

Education

Each of the risks and benefits outlined above will have implications for teachers of all subjects and at all educational levels.

The potential for AI to play a positive role in teaching and learning is clear. The ability to find and analyse information using natural language, to direct students by questioning their thought processes in a one-to-one or group setting, to create music and visual art, in language teaching: these and many more examples can all play a part in providing true individualised learning.

But the risks need to one taken very seriously if the bad is not to overwhelm the potential for good.

Systemic risks will, possibly, only involve teachers in making students aware of the limitations of systems like ChatGPT. Even in their current state (and they will improve), their output is convincingly human. While there is already a huge amount of dis- and mis-information readily available on the Internet that takes time to spot and debunk, it is highly likely "fact-checking" will become ever harder as AI-generated dis- and mis-information is distributed. Students need to be aware of this and be prepared to think critically about what they see and read online.

This needs considerable care as being a "critical-thinker" can easily turn to being a "cynical thinker" where all sources, however reputable, are distrusted.

While the ethics of the way AIs are created and trained is not something most teachers or students need to be concerned with, their use should be. For example, how much GAI help is acceptable? Maybe the AI can write the whole essay: but would it be ethical to submit this as your own work? Certainly, we would argue that not only is it not ethical, but it is also educationally damaging: this is particularly true when you cannot even trust that any facts quoted in the essay given are actually real (see "Hallucinations" above). Maybe this is the point that needs to be made: the use of a GAI in this way is equivalent to cheating, and you may still get the wrong answer.

The risks that GAI pose to society at large - through job losses, etc. - make interesting debating points in debating clubs and those classes that try to understand society. Indeed, some families may already be feeling the effects of AI on jobs: for example, BT is planning to cut 55,000 jobs over the new few years, of which 20% are a direct result of introducing AI.

The risk most likely to impact schools (both in admin and teaching/learning) comes from the "Bad-Actor" scenarios. Schools are already seen as "soft" targets for scammers and need to take great care in how they protect themselves from a multitude of cyber threats. The NEN offers advice on various topics including cybersecurity: Cybersecurity Guidance, a Cybersecurity Checklist and guidance on what to do if you have a cybersecurity incident.

GAI will likely be used by scammers to make their interactions with staff more plausible: by, for example, using a convincing, AI synthesised voice that can react more naturally during a scam telephone conversation. Students are likely to be bombarded with AI-generated content in their social media that will affect their worldview: the fact that boys are being affected by the misogynistic views of Andrew Tate even before deep-fake videos are common should serve as a warning of what may be to come as the software to create deep-fakes becomes more widespread and easier to use.

Teachers at all levels - Primary, Secondary and Tertiary - all need to try and understand how the new generation of AIs will impact them, their school, and what their students are being exposed to.

Governments are becoming aware of the issues and starting to think about both the risks and benefits of AI in general and GAI in particular. In March 2023, the UK's DfE released a paper on this: *Generative artificial intelligence in Education*. A paper published by the EU (September 2022) also considers the ethical use of AI in education. The EU paper is slightly out of date now as they do not take account of the most recent advances in Generative AI and Large Learning Models. It does, however, cover the previous generation AI models (ML, Natural Language Processing, etc.) to which many of the same risks and advantages apply. Perhaps the most recent and comprehensive paper was published in May 2023 by the US Office of Educational Technology entitled *Artificial Intelligence and the Future of Teaching and Learning*.

Existential Threat?

Is AI an existential threat? Much of the media coverage seems to think so. Geoffrey Hinton, one of the great pioneers of AI, has just left Google and, when talking about whether AI would destroy us in a recent Guardian article, said: "*I thought it would happen eventually, but we had plenty of time: 30 to 50 years. I don't think that any more. And I don't know any examples of more intelligent things being controlled by less intelligent things – at least, not since Biden got elected.*" He now thinks the crunch time will come in the next five to 20 years, he says. "*But I wouldn't rule out a year or two…*".

The idea of a (AI) Superintelligence that is so far above ours as we are from a frog is what is scary: if such an intelligence existed, how could we control it? This is a familiar trope in science fiction, from Arthur Clarke's HAL in "2001: A Space Odyssey", to Isaac Asimov's "Robot" series and his three laws of robotics.

In Nick Bostrom's book "Superintelligence" he tries to imagine what this would look like and what we should do now to protect ourselves. In his TED Talk he discusses machines more intelligent than humans.

But not everyone in the field is convinced that a superintelligence is possible and certainly not that the current models even offer a route to one based, as they are, on statistics, algorithms and vast quantities of data. This article, for example, attempts to counter Bostrom's arguments.

While there is disagreement about whether the current LLM based AI models will eventually lead to a super intelligence it is clear that AIs will continue to get better and performing tasks that have traditionally been considered the preserve of a human intelligence.

Whatever the reality, there is plenty of information on the internet if you are interested in these esoteric discussions (see Further Reading, below).

Summary

Artificial Intelligence, long a standard trope in Science Fiction, has hit the media headlines with the public release of two Generative AIs (GAI) - ChatGPT and Bard. While AI has been simmering away in the background since the 1970s with advances coming in fits and starts it has already made an impact on all our lives: mainly in the form of Expert Systems and, more recently, with Machine Learning algorithms.

Expert Systems are rule based: restricted to a narrow domain they rely on translating the knowledge and experience of a human expert into a set of rules to guide the user to logically consistent solutions. The most common examples being medical diagnosis where the AI narrows down the options by posing questions to the user which are based on answers to previous questions until an appropriate confidence level is reached and the systems provides the diagnosis. The key here is that an expert system is pure logic: the path from tail query to diagnosis is rule based. Nothing else.

Machine Learning introduces the concept of a Neural Network which can "learn" to answer a questions after being "trained" on a specific dataset. Again the domain of expertise if limited to the domain of the training dataset which is highly curated to offer examples for the ML to learn from: for example, satellite images for battlefield intelligence. While ML systems will provide an answer they do not explain how it was arrived at.

GAIs represent a step change in how AIs are created. Trained using vast amounts of data, typically scraped from the internet, and based on Large Language Models (LLMs) GAIs create naturalistic text as output from natural language input. Interacting with ChatGPT, for example, is like a chat with a human. You use natural language (ask a questions and it replies) and it remembers previous requests in the same session so it becomes a conversation, not just a sequence of queries.

There is undoubted potential for GAI in education: to guide, research, prompt. It could suggest lessons plans, be a party in a group discussion, translate languages, etc. But there are plenty of potential risks as well: cheating in homework, polluting the media with toxic content, better scams, job losses, and its tendency to just make things up.

Governments in the UK and US, and the EU are starting to respond: all have recently published papers and guidance on the use of AI/GAI in the teaching and learning sectors as well as AI more generally.

The mainstream media are, at the time of writing (May 2023), making much of the existential threat that AI may, or may not, pose to humanity. Some well respected AI researchers are also seeing this as a live issues. Others are more sanguine and suggest that the current, statistical models, would never to able to form a superintelligence capable of doing harm to humanity. But who knows when the next breakthrough may come: one that moves on from the purely statistical to more meaningful manipulation and a truly conscience machine?

References

Anderson, S.T. and C. (2017). *What AI is -- and isn't*. [online] www.ted.com. Available at: https://www.ted.com/talks/sebastian_thrun_and_chris_anderson_what_ai_is_and_isn_t [Accessed 30 May 2023].

Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016). *Machine Bias*. [online] ProPublica. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed 30 May 2023].

bard.google.com. (n.d.). *Bard*. [online] Available at: https://bard.google.com/ [Accessed 30 May 2023].

Boston Dynamics (2019). *Boston Dynamics is changing your idea of what robots can do. | Boston Dynamics*. [online] Bostondynamics.com. Available at: https://www.bostondynamics.com/ [Accessed 30 May 2023].

Boston Dynamics (2022). *Spot® - The Agile Mobile Robot*. [online] Boston Dynamics. Available at: https://www.bostondynamics.com/products/spot [Accessed 30 May 2023].

Bostrom, N. (2015). *What happens when our computers get smarter than we are?* [online] Ted.com. Available at: https://www.ted.com/talks/nick_bostrom_what_happens_when_our_computers_get_smarter_than_we_are [Accessed 30 May 2023].

Cardona, M.A., Rodríguez, R.J. and Ishmael, K. (2023). *Artificial Intelligence and the Future of Teaching and Learning Insights and Recommendations*. [online] US Department of Education. Available at: https://www2.ed.gov/documents/ai-report/ai-report.pdf [Accessed 31 May 2023].

Creative Commons (2019). *Creative Commons*. [online] Creative Commons. Available at: https://creativecommons.org/ [Accessed 30 May 2023].

Das, S. (2022). Inside the violent, misogynistic world of tiktok's new star, andrew tate. *The Guardian*. [online] 6 Aug. Available at: https://www.theguardian.com/technology/2022/aug/06/andrew-tate-violent-misogynistic-world-of-tiktok-new-star [Accessed 30 May 2023].

Directorate-General for Education, Y. (2022). *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators*. [online] *Publications Office of the European Union*. LU: Publications Office of the European Union. Available at: https://data.europa.eu/doi/10.2766/153756.

Dodd, V., Police, V.D. and correspondent, crime (2023). Police accused over use of facial recognition at King Charles's coronation. *The Guardian*. [online] 3 May. Available at: https://www.theguardian.com/uk-news/2023/may/03/metropolitan-police-live-facial-recognition-in-crowds-at-king-charles-coronation [Accessed 30 May 2023].

Generative Artificial Intelligence in Education. (2023). [online] Department of Education. Available at: https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education [Accessed 31 May 2023].

Gizmodo. (2023). *GOP Launches the Age of AI-Generated Attack Ads*. [online] Available at: https://gizmodo.com/youtube-biden-ai-political-ad-gop-stable-diffusion-1850373150 [Accessed 30 May 2023].

Goodreads. (n.d.). *Superintelligence: Paths, Dangers, Strategies*. [online] Available at: https://www.goodreads.com/en/book/show/20527133 [Accessed 30 May 2023].

Grammarly Spotlight: How We Use AI to Enhance Your Writing. (2018). *How We Use AI to Enhance Your Writing | Grammarly Spotlight*. [online] Available at: https://www.grammarly.com/blog/how-grammarly-uses-ai/ [Accessed 30 May 2023].

Hackett, J. (2019). *NEN Cyber Security Guidance for Schools*. [online] NEN - The Education Network. Available at: https://nen.gov.uk/advice/nen-cyber-security-guidance-for-schools/ [Accessed 31 May 2023].

Hackett, J. (2020a). *Cybersecurity – What If?* [online] NEN - The Education Network. Available at: https://nen.gov.uk/advice/cybersecurity-what-if/ [Accessed 31 May 2023].

Hackett, J. (2020b). *Cybersecurity Checklist*. [online] NEN - The Education Network. Available at: https://nen.gov.uk/advice/cybersecurity-checklist/ [Accessed 31 May 2023].

Hern, A. and @alexhern (2023). 'We've discovered the secret of immortality. The bad news is it's not for us': why the godfather of AI fears for humanity. *The Guardian*. [online] 5 May. Available at: https://www.theguardian.com/technology/2023/may/05/geoffrey-hinton-godfather-of-ai-fears-for-humanity [Accessed 30 May 2023].

Ho, K.K. (2023). *Getty Images Sues Stability AI Over Photos Used to Train Stable Diffusion Image Generator*. [online] ARTnews.com. Available at: https://www.artnews.com/art-news/news/getty-images-lawsuit-stability-ai-12-million-photos-copied-stabile-diffusion-1234656475/ [Accessed 30 May 2023].

https://www.facebook.com/swapankmanna (2023). *4 Key Differences Between ChatGPT and Bard*. [online] www.techgenyz.com. Available at: https://www.techgenyz.com/2023/04/19/key-differences-between-chatgpt-bard/ [Accessed 30 May 2023].

Juliani, A. (2023). *Large Language Models Don't 'Hallucinate'*. [online] Medium. Available at: https://betterprogramming.pub/large-language-models-dont-hallucinate-b9bdfa202edf [Accessed 30 May 2023].

Khan, S. (2023). *Sal Khan: How AI could save (not destroy) education*. [online] www.ted.com. Available at: https://www.ted.com/talks/sal_khan_how_ai_could_save_not_destroy_education [Accessed 25 May 2023].

Kuenssberg, L. (2023). AI creator on the risks, opportunities and how it may make humans 'boring'. *BBC News*. [online] 13 May. Available at: https://www.bbc.co.uk/news/uk-politics-65582386 [Accessed 30 May 2023].

Mahdawi, A. (2023). Will AI free us from drudgery – or leave us jobless and hungry? *The Guardian*. [online] 30 May. Available at: https://www.theguardian.com/commentisfree/2023/may/30/will-ai-free-us-from-drudgery-or-leave-us-jobless-and-hungry [Accessed 30 May 2023].

New Scientist. (2014). *New Scientist*. [online] Available at: https://www.newscientist.com/ [Accessed 30 May 2023].

OpenAI (2022). *Introducing ChatGPT*. [online] OpenAI. Available at: https://openai.com/blog/chatgpt [Accessed 30 May 2023].

OpenAI (2023). *ChatGPT*. [online] chat.openai.com. Available at: https://chat.openai.com/ [Accessed 30 May 2023].

openai.com. (n.d.). *DALL·E: Creating images from text*. [online] Available at: https://openai.com/research/dall-e [Accessed 30 May 2023].

Poniewozik, J. (2023). TV's War With the Robots Is Already Here. *The New York Times*. [online] 10 May. Available at: https://www.nytimes.com/2023/05/10/arts/television/writers-strike-artificial-intelligence.html [Accessed 30 May 2023].

stablediffusionweb.com. (n.d.). *Stable Diffusion Online*. [online] Available at: https://stablediffusionweb.com/#demo [Accessed 30 May 2023].

Sweney, M. (2023). BT to axe up to 55,000 jobs by 2030 as it pushes into AI. *The Guardian*. [online] 18 May. Available at: https://www.theguardian.com/business/2023/may/18/bt-cut-jobs-telecoms-group-workforce [Accessed 30 May 2023].

Vincent, J. (2023). *The swagged-out pope is an AI fake — and an early glimpse of a new reality*. [online] The Verge. Available at: https://www.theverge.com/2023/3/27/23657927/ai-pope-image-fake-midjourney-computer-generated-aesthetic [Accessed 30 May 2023].

Welborn, A. (2023). *ChatGPT and Fake Citations*. [online] Duke University Libraries Blogs. Available at: https://blogs.library.duke.edu/blog/2023/03/09/chatgpt-and-fake-citations/ [Accessed 30 May 2023].

Wikipedia Contributors (2019a). *Automation*. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Automation [Accessed 30 May 2023].

Wikipedia Contributors (2019b). *Chinese room*. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Chinese_room [Accessed 30 May 2023].

Wikipedia Contributors (2019c). *Clickbait*. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Clickbait [Accessed 30 May 2023].

Wikipedia Contributors (2019d). *John Searle*. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/John_Searle [Accessed 30 May 2023].

Wikipedia Contributors (2019e). *Robotics*. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Robotics [Accessed 30 May 2023].

Wikipedia Contributors (2019f). *Three Laws of Robotics*. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Three_Laws_of_Robotics [Accessed 30 May 2023].

www.aimyths.org. (n.d.). *Myth: Superintelligence is coming soon*. [online] Available at: https://www.aimyths.org/superintelligence-is-coming-soon [Accessed 30 May 2023].

www.aiva.ai. (n.d.). *AIVA - The AI composing emotional soundtrack music*. [online] Available at: https://www.aiva.ai/ [Accessed 30 May 2023].

www.bbc.co.uk. (n.d.). *Sunday with Laura Kuenssberg - Will It Be Hard to Make a Living for Years to Come?* [online] Available at: https://www.bbc.co.uk/iplayer/episode/m001m0bf/sunday-with-laura-kuenssberg-will-it-be-hard-to-make-a-living-for-years-to-come [Accessed 30 May 2023].

www.netflix.com. (n.d.). *Coded Bias | Netflix*. [online] Available at: https://www.netflix.com/gb/title/81328723 [Accessed 30 May 2023].

www.youtube.com. (n.d.). *'Deep Fakes' Are Becoming More Realistic Thanks To New Technology*. [online] Available at: https://youtu.be/CDMVaQOvtxU [Accessed 30 May 2023].

www.youtube.com. (n.d.). *Do You Love Me?* [online] Available at: https://www.youtube.com/watch?v=fn3KWM1kuAw [Accessed 30 May 2023].

Further Reading

Bias

Trenholm, R. (n.d.). *Eye-opening documentary Coded Bias, streaming on Netflix APril 5, faces racist technology*. [online] CNET. Available at: https://www.cnet.com/culture/entertainment/coded-bias-review-eye-opening-netflix-documentary-faces-up-to-racist-tech/ [Accessed 30 May 2023].

Education

Digital Learning Institute. (2022). *AI in Digital Learning*. [online] Available at: https://www.digitallearninginstitute.com/blog/ai-in-digital-learning-benefits-applications-and-challenges/ [Accessed 30 May 2023].

My students are using AI to cheat. Here's why it's a teachable moment. (2023). *The Guardian*. [online] 19 May. Available at: https://www.theguardian.com/technology/2023/may/18/ai-cheating-teaching-chatgpt-students-college-university [Accessed 30 May 2023].

UNESCO (2019). *Artificial intelligence in education | UNESCO*. [online] www.unesco.org. Available at: https://www.unesco.org/en/digital-education/artificial-intelligence [Accessed 30 May 2023].

Ethics

Harris, J. and @johnharris1969 (2023). 'There was all sorts of toxic behaviour': Timnit Gebru on her sacking by Google, AI's dangers and big tech's biases. *The Guardian*. [online] 22 May. Available at: https://www.theguardian.com/lifeandstyle/2023/may/22/there-was-all-sorts-of-toxic-behaviour-timnit-gebru-on-her-sacking-by-google-ais-dangers-and-big-techs-biases [Accessed 30 May 2023].

Merritt, R. (2023). *Moving Pictures: NVIDIA, Getty Images Collaborate on Generative AI*. [online] NVIDIA Blog. Available at: https://blogs.nvidia.com/blog/2023/03/21/generative-ai-getty-images/ [Accessed 30 May 2023].

Staff, B. (2019). *10 Best Deepfake Apps and Websites You Can Try for Fun*. [online] Beebom. Available at: https://beebom.com/best-deepfake-apps-websites/ [Accessed 30 May 2023].

Existential Threat

Cantor, M. (2023). This economist won every bet he made on the future. Then he tested ChatGPT. *The Guardian*. [online] 7 Apr. Available at: https://www.theguardian.com/technology/2023/apr/06/chatgpt-ai-bryan-caplan-interview [Accessed 31 May 2023].

Krier, S. (2022). *AI from Superintelligence to ChatGPT - Works in Progress*. [online] worksinprogress.co. Available at: https://worksinprogress.co/issue/ai-from-superintelligence-to-chatgpt [Accessed 31 May 2023].

Kuenssberg, L. (2023). AI creator on the risks, opportunities and how it may make humans 'boring'. *BBC News*. [online] 13 May. Available at: https://www.bbc.co.uk/news/uk-politics-65582386 [Accessed 31 May 2023].

Levy, S. (2017). *The Doomsday Invention*. [online] The New Yorker. Available at: https://www.newyorker.com/magazine/2015/11/23/doomsday-invention-artificial-intelligence-nick-bostrom [Accessed 31 May 2023].

Wikipedia Contributors (2019). *Superintelligence*. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Superintelligence [Accessed 31 May 2023].

Hallucinations

Everything, T.P. of, Shephard, A., Shephard, A. and Geoghegan, T. (2023). *The Great A.I. Hallucination*. [online] The New Republic. Available at: https://newrepublic.com/article/172454/great-ai-hallucination-chatgpt [Accessed 30 May 2023].

The Stack. (2023). *The Big Hallucination: Large Language Models and Consciousness*. [online] Available at: https://thestack.technology/the-big-hallucination-large-language-models-consciousness/ [Accessed 30 May 2023].

Wikipedia. (2023). *Hallucination (artificial intelligence)*. [online] Available at: https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence) [Accessed 30 May 2023].

www.unite.ai. (n.d.). *What Are LLM Hallucinations? Causes, Ethical Concern, & Prevention - Unite.AI*. [online] Available at: https://www.unite.ai/what-are-llm-hallucinations-causes-ethical-concern-prevention/ [Accessed 30 May 2023].

Misinformation

Cassauwers, T. (2019). *Can artificial intelligence help end fake news? | Research and Innovation*. [online] ec.europa.eu. Available at: https://ec.europa.eu/research-and-innovation/en/horizon-magazine/can-artificial-intelligence-help-end-fake-news [Accessed 30 May 2023].

Deepfakes web β. (n.d.). *Deepfakes web β | The best online faceswap app*. [online] Available at: https://deepfakesweb.com/ [Accessed 30 May 2023].

Reuters Institute for the Study of Journalism. (n.d.). *Will AI-generated images create a new crisis for fact-checkers? Experts are not so sure*. [online] Available at: https://reutersinstitute.politics.ox.ac.uk/news/will-ai-generated-images-create-new-crisis-fact-checkers-experts-are-not-so-sure [Accessed 30 May 2023].

Staff, B. (2019). *10 Best Deepfake Apps and Websites You Can Try for Fun*. [online] Beebom. Available at: https://beebom.com/best-deepfake-apps-websites/ [Accessed 30 May 2023].

Susarla, A. and Conversation, T. (n.d.). *How Artificial Intelligence Can Detect – and Create – Fake News | Hippo Reads*. [online] hipporeads.com. Available at: https://hipporeads.com/how-artificial-intelligence-can-detect-and-create-fake-news/ [Accessed 31 May 2023].

Taylor, J. and Hern, A. (2023). 'Godfather of AI' Geoffrey Hinton quits Google and warns over dangers of misinformation. *The Guardian*. [online] 2 May. Available at: https://www.theguardian.com/technology/2023/may/02/geoffrey-hinton-godfather-of-ai-quits-google-warns-dangers-of-machine-learning [Accessed 30 May 2023].

Villasenor, J. (2020). *How to deal with AI-enabled disinformation*. [online] Brookings. Available at: https://www.brookings.edu/research/how-to-deal-with-ai-enabled-disinformation/ [Accessed 30 May 2023].

Regulation

Artificial Intelligence and the Future of Teaching and Learning Insights and Recommendations. (2023). Available at: https://www2.ed.gov/documents/ai-report/ai-report.pdf [Accessed 31 May 2023].

Directorate-General for Education, Y. (2022). *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators*. [online] *Publications Office of the European Union*. LU: Publications Office of the European Union. Available at: https://data.europa.eu/doi/10.2766/153756 [Accessed 31 May 2023].

Generative artificial intelligence in education Departmental statement. (2023). Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1146540/Generative_artificial_intelligence_in_education_.pdf [Accessed 31 May 2023].

GOV.UK (2023). *A pro-innovation approach to AI regulation*. [online] GOV.UK. Available at: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper [Accessed 31 May 2023].

GOV.UK. (n.d.). *Office for Artificial Intelligence*. [online] Available at: https://www.gov.uk/government/organisations/office-for-artificial-intelligence [Accessed 31 May 2023].

Nast, C. (n.d.). *The Global Battle to Regulate AI Is Just Beginning*. [online] Wired UK. Available at: https://www.wired.co.uk/article/the-global-battle-to-regulate-ai-is-just-beginning [Accessed 31 May 2023].

TedTalks

Brockman, G. (2023). *Greg Brockman: The inside story of ChatGPT's astonishing potential*. [online] www.ted.com. Available at: https://www.ted.com/talks/greg_brockman_the_inside_story_of_chatgpt_s_astonishing_potential [Accessed 30 May 2023].

Chaudhri, I. (2023). *The disappearing computer -- and a world where you can take AI everywhere*. [online] www.ted.com. Available at: https://www.ted.com/talks/imran_chaudhri_the_disappearing_computer_and_a_world_where_you_can_take_ai_everywhere [Accessed 30 May 2023].

Marcus, G. (2023). *Gary Marcus: The urgent risks of runaway AI -- and what to do about them*. [online] www.ted.com. Available at: https://www.ted.com/talks/gary_marcus_the_urgent_risks_of_runaway_ai_and_what_to_do_about_them [Accessed 30 May 2023].

Russell, S. (2017). *3 principles for creating safer AI*. [online] www.ted.com. Available at: https://www.ted.com/talks/stuart_russell_3_principles_for_creating_safer_ai [Accessed 30 May 2023].

What is AI?

Klingler, N. (2021). *What Is an AI Model? Here's what you need to know*. [online] viso.ai. Available at: https://viso.ai/deep-learning/ml-ai-models/ [Accessed 30 May 2023].

McKinsey (2023). *What is ChatGPT, DALL-E, and generative AI? | McKinsey*. [online] www.mckinsey.com. Available at: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai [Accessed 30 May 2023].

Platform, E.L. (2021). *Artificial Intelligence For Dummies – A Guide For Beginners*. [online] Liberties.eu. Available at: https://www.liberties.eu/en/stories/artificial-intelligence-for-dummies/43527utm_medium=adgrants [Accessed 30 May 2023].

Wikipedia. (2023). *Generative artificial intelligence*. [online] Available at: https://en.wikipedia.org/wiki/Generative_artificial_intelligence [Accessed 30 May 2023].

Work

Boston Dynamics. (n.d.). *From Back-Breaking To Groundbreaking: The Future of Warehouse Automation*. [online] Available at: https://www.bostondynamics.com/resources/blog/back-breaking-groundbreaking-warehouse-automation [Accessed 31 May 2023].

English, S. (2023). *BT to shed up to 55,000 jobs by end of decade as robots take over*. [online] Evening Standard. Available at: https://www.standard.co.uk/business/bt-ai-job-cuts-philip-jansen-b1081948.html [Accessed 31 May 2023].

Kelly, P. (2023). 'Why would we employ people?' Experts on five ways AI will change work. *The Guardian*. [online] 12 May. Available at: https://www.theguardian.com/global-development/2023/may/12/why-would-we-employ-people-experts-on-five-ways-ai-will-change-work [Accessed 31 May 2023].

Naughton, J. (2023). Programmers, beware: ChatGPT has ruined your magic trick. *The Observer*. [online] 1 Apr. Available at: https://www.theguardian.com/commentisfree/2023/apr/01/chatgpt-write-code-computer-programmer-software [Accessed 31 May 2023]