# The Future of AI in Education:
# 13 things we can do to minimize the damage

**Arran Hamilton** (Cognition Learning Group)[1]

**Dylan Wiliam** (University College London)

**John Hattie** (University of Melbourne)

8 August 2023

## Abstract

We may already be in the era of 'peak humanity', a time where we have the greatest levels of education, reasoning, rationality, and creativity – spread out amongst the greatest number of us. A brilliant result of the massification of universal basic education and the power of the university.

But with the rapid advancement of Artificial Intelligence (AI) that can already replicate and even exceed many of our reasoning capabilities – there may soon be less incentive for us to learn and grow. The grave risk is that we then become de-educated and de-coupled from the driving seat to the future.

In all the hype about AI, we need to properly assess these risks to collectively decide whether the AI upsides are worth it and whether we should 'stick or twist'. This paper aims to catalyse the debate and reduce the probability that we sleepwalk to a destination that we don't want and can't reverse back out of. We also make 13 clear recommendations about how AI developments could be regulated - to slow things down a little and give time for informed choices about the best future for humanity. Those potential long-term futures include: (1) AI Curtailment; (2) Fake Work; (3) Transhumanism; and (4) Universal Basic Income – each with very different implications for the future of education.

**Additional Keywords and Phrases:** Future of Work; Purposes of Education; Artificial General Intelligence; AI Regulation; Deskilling of Humans; Brain-Computer Interfaces; Universal Basic Income; Fake Work; Transhumanism.

## Introduction

Throughout history, humans have fantasized about the possibility of creating thinking machines from inanimate mechanical parts. The ancient Greeks told mythical stories about Talos – a giant bronze automaton constructed by the god of blacksmiths. Leonardo De Vinci sketched drawings of humanoid robots; Isaac Asimov introduced a mechanical rogue villain in *I Robot* (1950); and in 1968

[1] Corresponding author email ahamilton@cognitionlearninggroup.com

Arthur C. Clarke showed the take-over power of the artificially intelligent HAL in *2001: A Space Odessey*, which was set 22 years in our past!

But all of this seemed like an utter fantasy, until 1956 when John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon came together for six weeks at the Dartmouth Conference to establish the field of Artificial Intelligence (AI). At the time, they were supremely optimistic that it would be possible to develop human-level thinking machines within a matter of years. Policymakers also took the prospect extremely seriously, with Lyndon B. Johnson establishing the US National Commission on Technology, Automation, and Economic Progress in 1964 (United States, 1966).

Yet despite these high aspirations, there was little progress until the late 1990s, and even after Deep Blue beat Gary Kasparov at chess in 1997, many remained sceptical that it would ever be possible to develop general artificial intelligence that would be able to think like us.

Within the last few years, however, opinions have changed, particularly with the arrival of Large Language Models. In November 2022, OpenAI launched ChatGPT-3 to the public, which did surprisingly well at many tasks that required higher-order thinking skills (and years of expensive education). ChatGPT-3 was quickly followed by GPT-3.5, and then by GPT-4 —each a noticeable improvement on its predecessor, and many other companies have developed their own Large Language Models.

In addition to the attention AI is now receiving in the mainstream media, it is also generating much interest in the world of education, teaching, and learning. Our concern, however, is that much of this is parochial. The focus is on issues in the here and now: tips and tricks to prompt the machines for better outputs; (reasonable) concerns about how to stop students using AI to cheat; and optimism that the technology can reduce teacher workload and act as a digital teaching aide for students. See Figure 1 for a summary of these near-view concerns.

**Figure 1: AI Benefits and Concerns for Education**

| Benefits | Concerns |
|---|---|
| Scaling quality education in hard to reach contexts; greater personalization; adaptive content; democratization of access; reductions in the cost of educational service delivery; teacher workload reductions; culturally relevant content; continuous assessment and feedback; coaching systems that help us to maintain goal commitment; decision support systems that help educational institutions to develop 'good' plans and stick to them; AI digital tutors instead of expensive high-intensity human tuition for learners with additional needs; AI support to teachers through augmented reality heads-up displays; deeper inferences about student learning through bio feedback tracking; faster identification and support for neuro-diverse learners; and multi-modal access for children with disabilities. | Accuracy (producing confident but incorrect answers); content bias; data protection, privacy and protection; plagiarism; surveillance of students and teachers; systems using poor pedagogic reasoning to increase pace of instruction for some students and to reduce it for others; equity of outcomes; algorithmic discrimination – where systems are used to make enrolment decisions/identify students for additional support; AI systems and tools that make decisions where it is impossible to understand how they are made; and worries that AI models take little account of context, and are focused on shallow factual knowledge. |

Source: author adaptation from US Department for Education Office of Educational Technology (2023); UNESCO (2021, 2022 & 2023); Giannini, (2023).

These are all significant concerns, but we think there are much more serious issues that are receiving too little attention:

## The future of education and learning!

Do schools and universities have a future, or will the machines soon be able to do everything we can – and better? Within the last decade, many educational policymakers were suggesting that we should teach students how to program, although there were others, such as Andreas Schleicher (the OECD's head of education) arguing that it was a waste of time, as machines would be as good as humans at such tasks (Schleicher, 2019). **But what if machines quickly advanced to a level where we could not even equal them or fully understand what they were doing in literally every domain?** Would this new world leave us with a profound motivation gap and the risk we become permanently de-skilled and de-educated?

Thankfully the future is not (yet) set in stone. There are many different possibilities. Some have us still firmly in the driving seat, leveraging our education and collective learnings. However, in some of the other – less positive – possible futures, humanity might lose its critical reasoning skills, because the advice of the machines is always so good, so oracle-like, that it becomes pointless to think for ourselves, or to learn.

But to discuss these issues we think it is helpful to begin with a short explanation of how the Large Language Models (LLMs) that dominate current work on artificial intelligence—ChatGPT, Google Bard, Claude, and Meta LLaMA —work; what they are capable of; how they are similar/different to human brains; and what the implications might be for human learning and motivations to learn. This is the focus of Part One of the paper. In Part Two, we explore four different scenarios for humanity and in particular, what each of these scenarios might mean for the future of learning and education. Finally, in Part Three, we present 13 recommendations that we think will help to ensure that AI becomes our greatest success, rather than a tangled mess. Figure 2 provides a summary of the paper.

**Figure 2: A Summary of the Paper**

| Waypoint | Key Point |
|---|---|
| ***Part One: The Discombobulating Background*** | |
| **1.1 How do our brains work?** | Our brains are largely about wiring and firing and this can be thought of as a kind of computational process. |
| **1.2 How do AI Large Language Models work?** | These systems direct attention, plan, reason and regulate themselves in ways that are similar to the way our brains carry out these functions. |
| **1.3 But aren't humans much more than machines?** | Many of our distinctly human capabilities, such as emotion, empathy, and creativity can be explained and modelled by algorithms, so that machines can increasingly pretend to be like us. |
| **1.4 What is the current capability of these AI Systems?** | These systems already exceed the 'average' human in a number of domains—perhaps a middling postgraduate but still prone to bouts of error. |

| | |
|---|---|
| **1.5 Have these AI systems reached their peak potential?** | No, we should expect them to greatly surpass human reasoning capabilities – possibly very rapidly; think thousands of geniuses thinking millions of things at once. Eventually their computational capacity is likely to exceed the combined brainpower of every human that has ever lived. |
| **1.6 What has happened to human skills during other eras of technological innovation?** | They were eroded, but this allowed us to focus on learning better/higher-order things. |
| **1.7 What are the Implications for human skills development the AI Era?** | Experts may—initially at least—see their capabilities considerably amplified, but novices could become forever stunted. We might even be living in the era of 'peak education'. As AI capabilities grow, our incentives to learn might diminish. It is not inconceivable that many of us might even lose the ability to read and write as these skills would, for many, serve no useful purpose in day-to-day living. |
| **Part TWO: Four Long-Term Scenarios**<br><br>*These scenarios speculate about potential futures for employment, education, and humanity – given what we have already unpacked in the discombobulating background of Part One.* | |
| **Scenario 1: AI is banned.** | Governments come together and ban future developments in AI.<br>---<br>We do not think this scenario is likely, but AI development could be slowed to ensure better regulation and safety, and to give time for careful consideration of which of the other three scenarios we want.<br><br>However, if future developments in AI were banned humans would still be in the driving seat and still require education. There might also be significant benefits for human learning from leveraging the AI systems that have been developed so far and that might, subject to satisfactory guardrails, be excluded from any ban. |
| **Scenario 2: AI and Humans work side-by-side (a.k.a. Fake Work).** | The AI gets so good that it can do most, if not all, human jobs. But governments legislate to force companies to keep humans in the labour market, to give us a reason to get up in the morning.<br>---<br>We think this scenario is possible in the medium-term – beyond 2035 – as Large Language Models and other forms of AI become ever more sophisticated.<br><br>But it may be highly dispiriting for humans to be 'in the room' whilst AI makes the decisions and no longer being at the forefront of ideas or decision-making. Even with increased education we would be unlikely to overcome this, or to think at the machines' speed and levels of sophistication. |

| | |
|---|---|
| **Scenario 3: Transhumanism where we upgrade our brains.** | We choose to upgrade ourselves through brain-computer interfaces to compete with the machines and remain in the driving seat.<br>---<br>We think this scenario is possible in the longer term – beyond 2045 – as brain-computer interfaces become less invasive and more sophisticated. But as we become more 'machine-like' in our thinking, we may be threatened with potentially losing our humanity in the process.<br><br>There would also no longer be any need for schooling or university, because we could 'download' new skills from the cloud. |
| **Scenario 4: Universal Basic Income (UBI).** | We decouple from the economy, leaving the machines to develop all the products and services; and to make all the big decisions. And we each received a monthly 'freedom dividend' to spend as we wish.<br>---<br>Community-level experiments in universal basic income (UBI) have already been undertaken and widespread adoption of this scenario could be possible by 2040. Some of the AI developers, including Sam Altman, are already advocating for this.<br><br>It would enable us to save our 'humanity' by rejecting digital implants but potentially we would have no reason to keep learning, with most of the knowledge work and innovation being undertaken by the machines. We might pass the time playing parlour games, hosting grand balls, or learning and executing elaborate rituals. We would perhaps also become ever more interested in art, music, dance, drama, and sport. |

## Where do we collectively go from here? 13 Recommendations

*These recommendations propose regulations to slow down the rate of AI advancement, so that we can collectively think and agree which of the four scenarios (or other destinations) suits humanity best. Otherwise, the decision will be taken for us, through happenstance and it may be almost impossible to reverse. We offer these as stimulus for further debate rather than as a final, definitive proposals but we believe that we need to conclude that debate FAST.*

1. We should **work on the assumption that we may be only two years away from Artificial General Intelligence (AGI)** that is capable of undertaking all complex human tasks to a higher standard than us and at a fraction of the cost. Even if AGI takes several decades to arrive, the incremental annual improvements are still likely to be both transformative and discombobulating.

2. Given these potentially short timelines, we need to quickly establish a **global regulatory framework**–including an international coordinating body and country-level regulators.

3. AI companies should go through **an organizational licensing process** before being permitted to develop and release systems 'into the wild' – much like the business/product licensing required of pharmaceutical, gun, car, and even food manufacturers.

4. **End-user applications should go through additional risk-based approvals** before being accessible to members of the public, similar to what pharmaceutical companies need to do to get drugs licensed. These processes should be proportionate with the level or risk/harm – with applications involving children, vulnerable or marginalized people being subject to much more intensive scrutiny.

5. **Students (particularly children) should not have unfettered access to these systems before risk-based assessments**/trials have been completed.

6. **Systems used by students should always have "guardrails" in place that enable parents and educational institutions to audit how and where children are using AI in their learning.** For example, this could require permission from parents and school prior to being able to access AI systems.

7. Legislation should be enacted to make it **illegal for AI systems to impersonate humans** or for them to interact with humans without disclosing that they are an AI.

8. **Measures to mitigate bias and discrimination in AI systems should be implemented**. This could include guidelines for diverse and representative data collection and fairness audits during the LLM development and training process.

9. Stringent **regulations around data privacy and consent**, especially considering the vast amounts of data used by AI systems. The regulations should define who can access data, under what circumstances, and how it can be used.

10. **Require AI systems to provide explanations for their decisions** wherever possible, particularly for high-stakes applications like student placement, healthcare, credit scoring, or law enforcement. This would improve trust and allow for better scrutiny and accountability.

11. As many countries are now doing with the Internet systems, **distributors should be made responsible for removing untruths, malicious accusations, and libel claims** – and within a very short time of being notified.

12. **Establish evaluation systems to continuously monitor and assess the safety, performance, and impact of AI applications**. The results should be used to update and refine regulations accordingly and could also be used by developers to improve the quality and usefulness of their applications – including for children's learning.

13. **Implement proportionate penalties for any breach of the AI regulations**. The focus could be creating a culture of responsibility and accountability within the AI industry and end-users.

Again, there will be differences of opinion about some of these – so you can treat them more as stimulus to further debate, rather than as a final set of cast-iron proposals. But we need to have that debate FAST and then enact pragmatic measures that give us breathing room to decide what kind of future we want for humanity – before it is simply foisted upon us.

Before we scare you too much, however, with our four scenarios – as noted above, we do not believe that the future is pre-determined. There are many other possible outcomes with different (happier) endings. However, we think we all need to understand the dystopian possibilities before we accidentally venture down a path with no escape route. This is more important than ever as we are arguably at what Will MacAskill (2022) calls 'the hinge point of history' i.e., that moment where things accelerate faster than ever before, where things move, for example, from linear to exponential. It is this that motivates our recommendations, helping to slow down the rate of progress and collectively giving us time to think.

Some bullish researchers already think we may only be two years out from Artificial General Intelligence that can reason to the same standard as you or us (Cotra, 2023). Most others – the bearish – still think it will likely be with us before 2040 i.e., the time at which today's toddlers graduate from high-school; and quite possibly sooner (Rosser, 2023).

Our own position is that there is great uncertainty but that we ALL need to maintain a stance of vigilance and assume – from now on – that at any moment in time we *could* be only two years out from machines that are *at least* as capable as us. So, we can't bury our heads in the sand or get all parochial – we need to grapple with these ideas and their implications today.

# *Part One: The Discombobulating Background*

In this section we review how AI is similar/different to us; where the tech is going; the implications for human skills development, and of course for education! And we start by recapping some of the key features of the human brain drawing out the parallels with AI systems.

## 1.1 How do our brains work?

To the best of our knowledge, the human brain is *currently* the most sophisticated computational device in the known universe. Packed within its grey meaty matter are an average of 83 billion neurones (Herculano-Houzel, 2009), that flexibly "fire and wire" as we interact with our environment.

Some of the key features of this 'meat computer' are as follows (Eagleman, 2015; Seung, 2012; Damasio, 2000; O'Shea, 2005):

- **Specialist neurones** that have expertise in different processes – including vision, hearing, language, facial recognition, and abstract reasoning

- A **layered structure –** within each of these specialized domains there is a tightly packed hierarchy of neurons. Those at the bottom do more basic things, with ever growing complexity of computation occurring by the time the outputs reach the top

- **Flexible connections** – each neuron has protruding 'wires' (i.e., axons and dendrites) that connect with and pass information to other neurons within the network

- **Load weighting** – each neuron has a threshold for the number of signals that it needs to receive from other neurons for it to fire in turn and thus pass its signal on.

One way to think of this is as a kind of giant voting machine – with each layer comprising a group of judges that vote on ever more complex matters. When light hits your eyes, for example, a layer of neurons will judge whether photons are exciting the rods and cones in specific areas; and send their

votes upwards to the next layer in your visual cortex. These higher layers, in turn, will judge shapes, colours, textures and see, for example, a golden retriever. They will pass that information to other areas of the brain that, in turn, vote on whether the dog is friendly or dangerous. And yet others, which decide whether to pet it or run away.

This process is recursive – can you, for example, remember a time when you spied a dog or cat out of the corner of your eye but when you looked more closely it turned out to be a bag of rubbish or a fire hydrant? Your neurones were 'voting', calculating, and making probability inferences based on the information they had access to at the time and as the picture become clearer – they revised their votes, with all this happening in mere milliseconds. Your 'voting machine' is arguably engaging in a very sophisticated form of computation.

***In a Nutshell:*** **Our brains are largely about firing and wiring; and this can be thought of—or at least modelled as—a computational process.**

## 1.2 How do AI Large Language Models work?

Current AI large language models such as ChatGPT, Bard, Claude, and LLaMA are a type of deep learning neural network that operate remarkably similarly to aspects of the human brain. They contain (Radford et al, 2019; Brown et al, 2020; Bowman, 2023):

- **Parameters** – which are akin to digital synapses. ChatGPT-3 has an officially reported 175 billion parameters, and GPT-4 is considered to be considerably bigger, possibly as many as 1 trillion**.**

- **Hierarchical architecture** - with the input layers passing their votes to higher levels within the system**.**

- **Flexible connections and load weightings** – much like the human brain, the individual parameters within these models can flexibly adjust the 'digital synapses' at higher levels within the model to which they pass their outputs. They can also adjust the number of inputs or votes that they need to receive from lower-level parameters before they fire in agreement.

These models are trained on vast amounts of data – effectively the whole of the internet. When you ask one of these systems a question it breaks down the input text into smaller pieces called "tokens". It then reads these tokens one at a time to determine what it is being asked to do and to decide which words in the input text are most important and what they mean. Then, it produces an output response – selecting one word at a time i.e., by predicting what the best next word would be, laying this word down on the screen and repeating the process again and again – until it has produced hundreds or thousands of words of (usually) high-quality text. These models also "learn" – using feedback loops from their training, and reward tokens, to adjust their connections and load weightings to generate better outcomes.

Note, too, that the keystone paper from which the original idea of Large Language Models was inspired – came from early scientific research into human neurons (McCulloch & Pitts, 1943). Warren McCulloch was a neurophysiologist and psychiatrist who, along with logician Walter Pitts, was

seeking to build a mathematical model of the human brain. A version of this model is what drives these current AI systems.

> ***In a Nutshell:*** **these systems are not dissimilar to the human brain's prefrontal cortex—the part of us that does abstract thinking and that separates us from the animals.**

## 1.3 But aren't we humans much more than mere machines?

Often when people are confronted with the fact that our brains can be modelled as just computational devices and that large language models digitally mimic *some* of these functions, there is a strong desire to push back—to say that we are more than mere machines. Some of the most common objections are to say that "we have consciousness", "we have emotions", "we have goals", "we are creative", "we have empathy" and "they don't think like us". Basically, that we are smart, and they are toasters.

We want to, briefly, offer some contrasting perspectives on these claims:

- ***"We have Consciousness"***. Whilst this seems intuitively reasonable to us all, many philosophers and neuroscientists struggle to operationalize the concept at a subjective level (Dennett, 1991; 2005; Metzinger, 2009; Churchland, 2013). Some suggest that consciousness is "user illusion" – a kind of dashboard or user-interface that simplifies the complex computational outputs of the brain into a series of simpler heuristics to enable us to make quick decisions (Nørretranders, 1991). You could think of this as a kind of graphical user interface. Yet others think consciousness is just what it subjectively "feels like" to process information and that even atoms, thermostats, and computer chips experience consciousness (Chalmers, 2010; 2022; S. Harris, 2020; A. Harris, 2019).

- ***"We have emotions".*** One jarring interpretation is that our emotions are just electrical signals that operate as part of a computational decision-making process (Minsky, 2006; Prinz, 2004; 2017; Damasio, 2006; Barrett, 2017); that our brains process vast amounts of data – much of it subconsciously; and that many brain regions struggle to directly talk to neighbouring regions in the same 'programming language'. Instead, they send (a more informationally efficient) jolt to our stomach, increase our heart rates, or make us quiver with anxiety as a quick (but indirect) mechanism for those other brain regions to notice and respond to. These, in turn, whirr into action and seek to interpret what that 'gut feeling' or 'shiver down the spine' might mean.

- ***"We have goals".*** Unlike our deliberately jarring provocations to the two perspectives above, we humans do indeed have agency. We are able to pause, think, make long-term plans, reflect on those plans, make choices, and recalibrate our responses. But developing and evaluating goals is inherently algorithmic. We search through options, weigh the pros and cons, and develop a plan of action. We see no reason why future Artificial Intelligence systems could not be trained to have these recursive feedback loops and then to operate autonomously; to be agents, like us. (See Bostrom, 2014; and Russell & Norvig, 2016).

- ***"We are creative".*** Indeed, many of us are. But, (again) there are strong arguments that creativity can be modelled as an algorithmic process where we: (1) select a problem/goal space; (2) map current approaches; (3) identify alternatives; (4) pick one or more alternative to try; (5) test whether it is better than what we are currently doing; (6) decide where to next (see Boden,

2006; and Colton, 2008). Arthur Koestler (1964) defined creativity as the bringing together of many seemingly unrelated ideas. There is no reason why Artificial Intelligence could not do the same thing. Indeed, several COVID-19 vaccines were co-'invented' by a protein folding AI that went through a version of this creative process and proposed an optimal design quickly (see for example Sharma et al, 2022). New types of antibiotics are also being "creatively" invented in this way (Lluka & Stokes, 2023).

- ***"We have empathy"***. Yes, again, many of us do. One interpretation of this is that we are (computationally) simulating how another "meat computer" will process and respond to similar inputs and adjusting our approach in light of this understanding. Given that Artificial Intelligence systems do not mirror the design features of the human brain, it would likely be difficult for them to simulate, at a subjective level, human thoughts, and feelings or to see things through our eyes. But this might not matter.

  These systems can already use personality profiling tools and biometrics to predict our behaviour (see for example, Hoppe et al, 2018, on how AI can predict personality through human eye movement). If their goal is to influence us, all they need to do is to draw on these insights, try "talking to us" in a specific way, reflect on whether this resulted in the desired changes in our behaviour; and repeat until they find the right influencing strategy – for us. This can be thought of as the illusion of empathy but one that is possibly indistinguishable from actual empathy—arguably very similar to that employed by human sociopaths, that also seem to lack empathy circuitry.

- ***"They don't think like us"***. Again, we agree – AI systems almost certainly don't. They are currently most analogous to our pre-frontal cortex. But they don't use neurotransmitters like dopamine or serotonin, have no physical bodies that give out electrical jolts that can be interpreted as emotions, and have no hormone emitting glands. However, we do not think that means they are incapable of thought. Instead, many researchers in the AI field suggest that a good metaphor is to think of these systems as a form of 'Alien Intelligence' (see for example Rees & Livio, 2023) – that these machines think and process information, just in a different way to us. Those differences might give us advantages at certain tasks (we are a lot more energy efficient), but it might equally confer similar advantages to them in other areas – like analysing large datasets, writing high-quality prose, and developing complex strategies in seconds.

The perspectives highlighted above are not definitive. We have raised them because we all need to be shaken out of our human-centric view on intelligence and thought and to consider the possibility that other forms of highly intelligent life are possible. Max Tegmark (2017), the MIT Physicist, uses the term "Carbon Chauvinism" to illustrate this. His point is that complex computation can be undertaken in a variety of substrates/materials.

Our human substrate is carbon – the base element for the protein from which our "meat computers" are built. But Tegmark's suggestion is that other substrates like silicon could think and process information in exactly the same way. If the silicon neurons and their load weightings perform similar to ours, we could even create digital copies of ourselves.

At risk of going down a rabbit hole, Roger Penrose (1989) – the acclaimed the mathematical physicist – even has a thought experiment where he asks us to imagine that the biological neurons in our

brain are gradually replaced with identical silicone ones. They question is at what point do we stop being 'us'? There is the distinct possibility that we might not even notice the difference.

***In a Nutshell:*** **Our brains *might* be entirely computational, although, embracing our computational nature does not diminish our uniqueness. Instead, it highlights the power and sophistication of the computational processes within us!**

## 1.4 What is the current capability of these AI Systems?

Current systems have already advanced at a phenomenal rate during the last half-year alone. In Figure three below we tabulate human performance in a range of standardized academic tests, as compared to GPT-3.5 and GTP-4, both released in 2023.

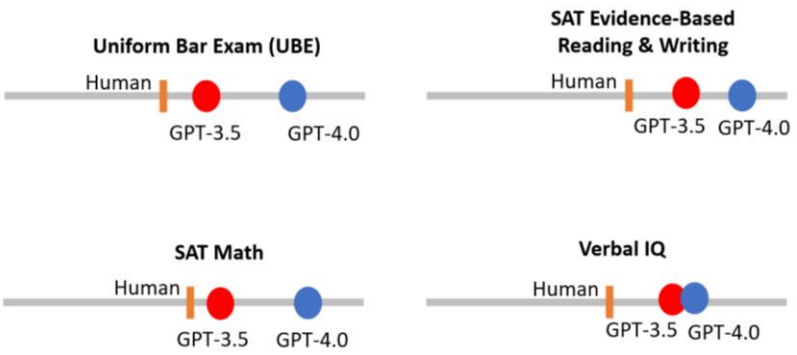
**Figure 3: Human vs ChatGPT on a Range of Tests**

| Test | Average Human | GPT-3.5 | GPT-4 | Maximum possible score |
|---|---|---|---|---|
| Uniform Bar Exam (UBE) | 140.3 | 213 | 298 | 400 |
| LSAT | 152 | 149 | 163 | 180 |
| SAT Evidence-Based Reading & Writing | 529 | 670 | 710 | 800 |
| SAT Math | 521 | 590 | 700 | 800 |
| Graduate Record Examination (GRE) Quantitative | 157 | 147 | 163 | 170 |
| Graduate Record Examination (GRE) Verbal | 151 | 154 | 169 | 170 |
| Verbal IQ | 100 | 147 | 155 | |

Source: OpenAI (2023); de Souza, de Andrade Neto & Roazzi (2023); and publicly reported human test taker scores in most recent test cycle

To put this into context, this means that GPT-4 already scores higher than 90% of humans on the Uniform Bar Exam and the SAT, 88% of Humans of the LSAT, and has a verbal IQ well within the top 0.02% of the human population. We graphically illustrate this in Figure four, below.

**Figure Four: Scaled Representation of Human vs ChatGPT on a Sub-set of Tests**




Source: Authors' graphical representation of selected data from OpenAI Technical Report (2023); de Souza, de Andrade Neto & Roazzi (2023); and publicly reported average human test taker scores

ChatGPT-4 has also passed the US Medical Licensing Examination; and achieved above average scores on selected Wharton MBA examinations (Terwiesch, 2023). And as can be seen in Figures three and four, the grades achieved on many forms of test are on a strong upward trajectory.

Now it is important to note that these systems are prone to hallucination or confabulation – making up convincing but incorrect or misleading answers – but so are we humans, when we are asked for off-the-cuff opinions on the fly. But much like humans, when you ask these systems to reflect on their first drafts – to check that the evidence provided is real and to improve what they have written – their outputs get better and better.

What is most impressive about these AI systems is their speed of response and ability to reflect on the output and improve it, without getting upset, when provided with the right kind of prompts. There are several organizations—both within education and the corporate world—undertaking work-productivity trials. They are experimenting with these systems to automate complex thinking tasks and there are anecdotal reports of a sizable increase in the productivity of content writers, coders, illustrators, and analysts – with the machines able to do the hard (and creative) grind and with humans checking and enhancing the outputs. In other words, humans working with machines.

In the same manner, computers have been as accurate as humans at scoring student essays for many years now, while also providing feedback that is as good as that provided by humans, and that is more acceptable to students because the comments are not biased by personal relationship issues. With the new options in AI, students could use the machine feedback to learn how to better write essays, have them marked immediately, and teachers could spend more time teaching, moderating, and helping students know how to go deeper, wider, and explore.

This also echoes earlier waves of AI deployment—where human chess players have for some time been using specialist chess engines for some time to perfect their technique to compete against other human players. It is also worth noting that the best (current) chess players are not humans nor machines, but teams of humans working with machines to figure out the best move—what is called "freestyle chess".

## Yes, but aren't the Machines just Mindless Plagiarists?

One of the common criticisms of the outputs of Large Language Models is that the content is all 'plagiarised' from the existing canon of work produced by other (human) authors. Of course, the standard definition of plagiarism is to make a verbatim copy of something without improvement or variation, and without attribution. You will remember back to your school/university days when your tutors instructed you to summarize people's ideas in *your own words* – rather than copy and paste.

Yet, large Language Models are rarely 'plagiarising' in this way. Instead, they are analysing an existing body of knowledge – whether that be written words, art, or music and drawing on this as inspiration to execute whatever task we ask of them, whether that be:

- To produce a wedding speech in the style of William Shakespeare
- To draw a picture of Mother Teresa riding a bicycle in the style of Salvador Dali
- To write the lyrics for a song about the benefits of fish and chips in the style of the Beatles.

They then use their training data, which includes the complete works of Shakespeare, Dali, and the Beatles – to execute our request. And what comes back is not usually a hack job that cuts and pastes pre-existing text, art, or music lyrics. It's brand-new stuff – which is why university plagiarism detection software has such a hard time uncovering it.

If this is considered plagiarism, then Shakespeare was also a plagiarist. After all, he drew on and extended the literary conventions of the day. Ditto Albert Einstein who built atop the works of Isac Newton, James Clark Maxwell, Henri Poincare, and many others. If Einstein had been born in a cave 50,000 years ago, there is literally zero chance that he would have stumbled across the theory of general relativity – because he would not have been able to 'plagiarise' the work of others. Except of course, he was not actually plagiarising. He was synthesizing, extending, and enhancing an existing body of knowledge and he reiterated Newton's claim that he was standing on the shoulder of giants – which is exactly what Large Language Models are doing.

We also see no reason why AI would not be able to build on top of its own work – creating synthetic outputs – in the way that we humans critique and extend one another's thinking. There could even be different AI systems that work in direct competition with one another to identify appropriate goals, seed hypothesis, establish plans and debate with one another – each being given a 'reward token' each time it wins the competition. This would be no different from contemporary human research and development - where people publish, critique, apply for patents and get promoted on the basis of publications.

***In a Nutshell:*** **These systems already exceed the 'average' human in a number of domains, sometimes by leaps and bounds; and they have accelerated fast in their capabilities over the past half-decade. They are not just 'dumb plagiarists.'**

## 1.5 Have these AI systems reached their peak potential?

In recent months, a number of important educational coordinating institutions have written reports on the implications of Artificial Intelligence for Education – including the US Department of Education (2023) and UNESCO (2023). One thing we have noticed is that these reports have focused on the implication of *current* AI systems on the *current* schooling system. The subtext in this thinking is – perhaps – that is unwise to speculate on future advances and timelines, because the future is a foreign country.

However, within the AI community a distinction is commonly made between different types of Artificial Intelligence (Bostrom, 2014; Ford, 2015; Kurtzweil, 2013; Norvig & Russell, 2021), illustrated in Figure 5.

**Figure 5: The Different Types of AI Capabilities**

| Type | Description | Examples of capabilities |
|---|---|---|
| **Narrow AI (i.e., One trick ponies)** | Domain specific systems that are extremely good at ONE specific thing | Exceed human-level skills in ONE domain e.g., Chess playing; Protein Folding; Stock Market Analysis; Image Creation; Translation etc. |
| **Artificial General Intelligence (AGI) – a.k.a. human-level intelligence** | Cross-domain systems that can operate at human-levels of ability in ALL cognitively demanding tasks | As good as humans in ALL tasks that require strategic thinking and analysis e.g., macro-economics; law; medicine; policy analysis; research; writing; decision-making; |

| | | |
|---|---|---|
| | | persuasion; coaching; computer programming etc. |
| **Artificial Super Intelligence (ASI)** | Cross-domain systems that operate well in advance of human-level in ALL cognitively demanding tasks | Ray Kurtzweil suggests that such systems could have greater cognitive abilities than the combined might of all humans currently alive and those that have ever lived. He calls this scenario "God in a box" (Kurtzweil, 2013; 2005) |

The current consensus in the AI community is that existing Large Language Models like GPT-4 already exceed the capabilities of "Narrow AI" and that they *might* already be exhibiting *some* sparks of "Artificial General Intelligence" (Bubeck et al, 2023). We also think that they already pass the Turing Test – proposed by the Alan Turing (1950) – where the threshold is simply that machine dialogue is indistinguishable from that of a human, and where machines can easily pass themselves off as people.

Figure 6, which was cross-tabulated by Charlie Giattino and Max Rosser at *Our World in Data*, illustrates the changes in expert opinion on the likely timelines for the emergence of AGI. This brings together opinions from surveys of industry experts (Grace et al., 2022; Zhang et al., 2019; & Gruetzemacher et al., 2018); crowdsourced opinions from professional forecasters (Meaculus, 2022); and the findings of an in-depth review by the thinktank Open Philanthropy (Cotra, 2023).

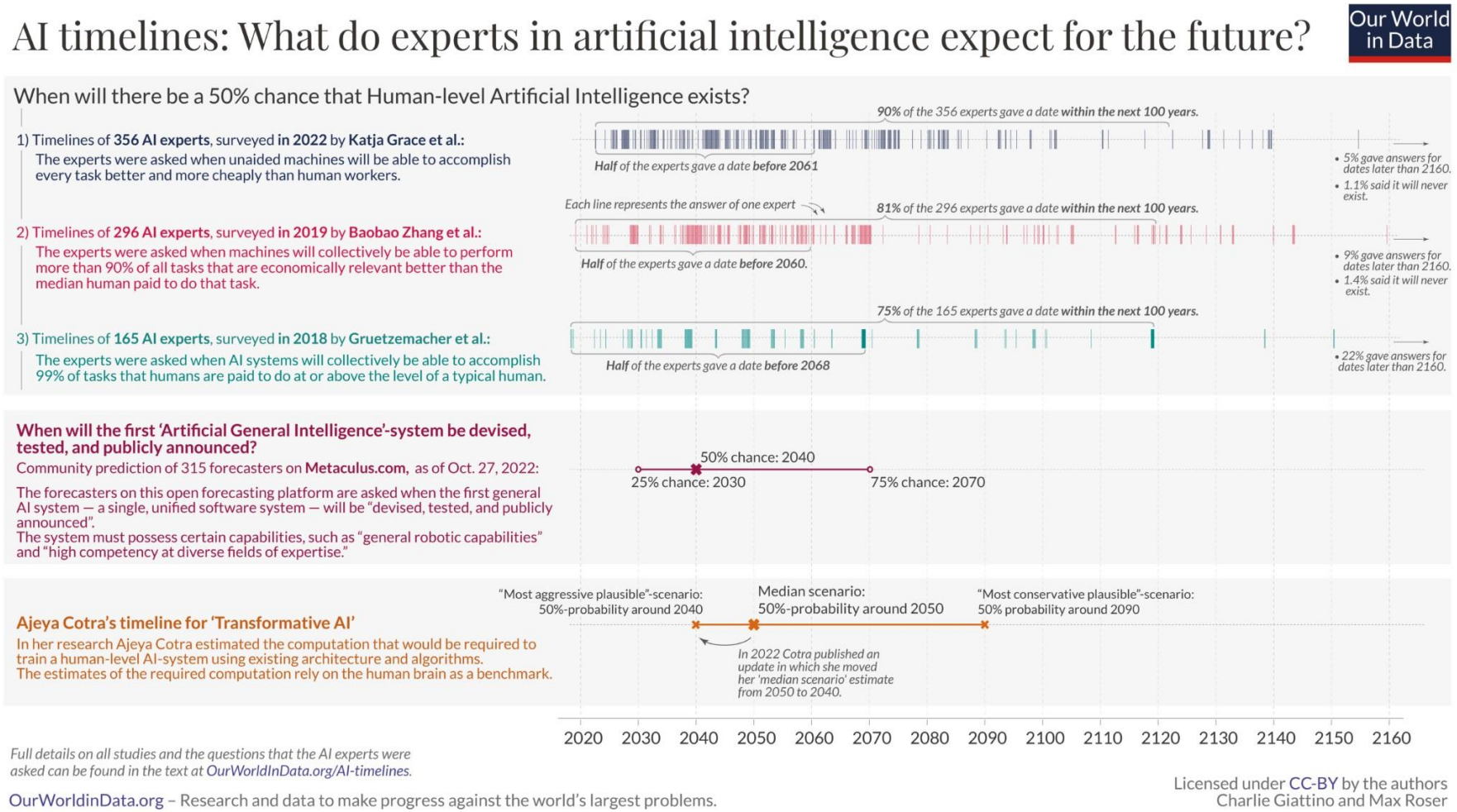
Our interpretation of these forecasts is that:

1. **Most experts in the AI field believe that it is merely a matter of time before these systems can achieve full 'Artificial General Intelligence'.** Indeed, less than 2% of recently surveyed experts thought that full AGI was a technical impossibility that could never be achieved.

2. **The experts also now seem to be revising their timelines downwards – with the expectation that AGI could arrive much earlier than their original estimates.** For example, Ajaya Cotra (2023) at Open Philanthropy has recently knocked 10 years off even her most conservative scenario. She now puts the 'median arrival date' at 2040 rather than 2050 and she does not discount the possibility of a major breakthrough by 2025.

3. **Whilst there is high confidence that AGI is possible and probable, there is still great uncertainty about the timelines for its arrival (ranging from less than two years to more than a century).** This mirrors the challenges that people at the turn of the 20th Century had in predicting when humans would conquer the skies (Roser, 2023). Wilbur Wright is quoted as saying "I confess that in 1901, I said to my brother Orville that man would not fly for 50 years." Yet, two years later powered flight had been achieved and by the very aviation experts that had apparently been so conservative in their estimates, just 24 months before! But we also need to remember the extreme overconfidence of the original Dartmouth AI Conference participants in the 1950s, who thought AI would be cracked in a few years.

**Figure 6: Artificial General Intelligence Timelines**



4. **Some of the surveyed experts in the AI field believe that it is possible that AGI will be achieved by 2030 and possibly even in the next two years.** They hold this view because the Large Language Models that drive ChatGPT are considered relatively simple and explainable to even middle-school students; they contain only a few thousand lines of code vs the 45 million lines in Microsoft Word, and so, it may require only a few additional – but equally simple steps – to achieve AGI. Accelerating this possibility is the fact the Large Language Models are increasingly able to program (and re-program) themselves—what we might call 'learning'. And, of course, parallel advances in the computational hardware that these systems run on, also speed up the rate of computation and advancement.

5. **Given all the above, we suggest a stance of hypervigilance and think it would be wise to continuously assume that *we could be only 2 years out from AGI*.**

But even if we are talking about AGI being 20 years out, this still has profound implications for the here and now. It is now common for education systems to proclaim that they are preparing young people for a 'world we can't imagine'—a world that is significantly different from that of today; that requires different (and ever more advanced) capabilities to survive and thrive. If it is a world where AI can undertake most or all the cognitively demanding work – the stakes for education might actually become lower.

The sense within the AI community, although not uniform, is that machines with Artificial General Intelligence could also quickly make the leap to Artificial Super Intelligence – because they would continuously and expertly re-program themselves to lock-in and extend newly acquired skills (Bostrom, 2014; Kurtzweil, 2013; Yudkowsky, 2008; Armstrong, et al., 2012; Tegmark, 2017). Some researchers think this could happen in a matter of weeks (fast take-off) and other suggest less than 5 years after AGI is achieved (slow take-off).

## Disembodied Systems

However, we think these systems are likely to be largely disembodied for at least the next decade, no matter how smart they are, because making advances in computational thinking has proven far easier than training AI to drive cars, walk-up stairs, or to peel an orange. As Hans Moravec pointed out 35 years ago:

> *"it is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility." (Moravec 1988 p. 15).*

Evolutionary biologists suggest that our physical and perceptual capabilities evolved over several hundred million years (see for example Lieberman, 2013; Heyes, 2012), whereas abstract thought and language likely only emerged in the last 200,000 years and may operate through the same relatively simple architecture as ChatGPT.

***In a Nutshell:*** **At some point – in the next two to twenty years – AI systems are likely to meet and then exceed human capabilities in many, most, or all cognitively demanding tasks, although it is likely to take considerably longer before they can match our physical abilities to shape and reshape the world.**

## 1.6 What has happened to human skills during other eras of technological innovation?

In previous waves of disruptive technological innovation there has often been a tug of war between two perspectives:

1. **Those fearful of the consequences of change (a.k.a. the Cassandras)** – think here of Plato who was worried about the impact that the invention of writing would have on human memory and cognition (Plato & Jowett, 2010); and the more radical 19th century Luddites that destroyed the textile machines that eroded the value of their skills and drove their wages down.

2. **Those that think *'it will all come out in the wash'*** – here, you can think of the economists and TED-talk gurus that look back at previous waves of disruptive innovation and that argue that whilst it was discombobulating that the time – the net benefits of these changes to life expectancy and standards of living have been vast (see for example Brynjolfsson, & McAfee, 2016; Romer, 1994; Solow, 1956; and Schumpeter, 2015). Further, most of the displaced workers moved from farms, to factories, to knowledge work – with the support of state-financed universal education.

In recent years, that second '*it will all come out in the wash*' perspective has become the dominant paradigm. The argument is that, yes, there has been a shedding of some skills that are no longer necessary but there is also collective climbing of the skills ladder to acquire new and more complex skills that are needed for the modern knowledge-economy.

The suggestion is that – historically – the new technologies have *always* resulted in new products, new industries, more and better jobs, greater productivity, and cheaper cost of goods and services; and that this is, therefore, is an 'Iron Law' of economics that will continue to be the case. The Cassandras have always turned out to be wrong in the past and thus it is likely that any AI-Cassandras must therefore also be wrong about the future, too.

The economists are correct that many skills have, indeed, been rendered less valuable as a result of previous waves of innovation:

- **Farming** – 300 years ago, an average of 70% of the global population worked in farming and this is now less than 2% i.e., a shift from 70 out of every 100 people to 2 out of every hundred (Lowder, Skoet, & Raney, 2016; Our World in Data, 2023a). Machines do the rest and agricultural output is higher than ever.

- **Mental arithmetic** – prior to the 1967 invention of the pocket calculator, people working in retail and many other fields needed strong mental arithmetic skills to quickly add up customer bills and give back the right change. This is no longer required and rarely taught in schools today, with the result that when an electronic till breaks down the cashier will often resort to their smartphone to do simple mental computations that used to be taught in primary school (see Ellington, 2000; 2006; and Hembree & Dessart, 1986 for an overview of the academic literature). The research literature suggests that whilst students' mental arithmetic skills may have declined through use of calculators, their minds are freed up to work on higher-order problems – exactly as the economists suggest.

- **Map reading** – the widespread availability of mapping apps on smartphones has undoubtedly helped those who found using printed maps challenging, but there is also evidence that these tools have reducing the accuracy of human mental maps and wayfinding (Shikawa et al., 2008; Gardony et al., 2013; and Sparrow, Liu, & Wegner, 2011).

Shedding these skills (and many others) is almost akin to having a personal assistant that does all the dirty work, so that you can instead focus on the bigger things. The time we have saved in not growing our own food, not grooming horses, not needing the check whether the cashier has added up properly, not needing plan our travel routes in advance, and not needing to rote memorize new phone numbers and addresses – frees time up to do better things, like thinking about the implications of AI for education and humanity; and writing this paper!

This skill-shedding has been a feature throughout human history, exactly as the economists have argued. Our ancestors needed to be good at hunting, gathering, tracking, self-defence, shelter-building and all the rest. Their lives depended on it. Ours do not, so most of us have *completely* shed those skills and climbed the cognitive ladder.

So far so good. But with the invention and continued advance of AI – this trend may not continue in quite the same way. **If the machines can do ALL the higher-order thinking tasks faster and more accurately than humans – and be super-Einstein, super-Eddison, super-Freud, and super-Abraham Lincoln, all rolled into one – there may be no 'next level up' for us to pivot to.** Instead, we may be left only with the niche areas that the machines can't (yet) do – many of which are back down at the lower levels of the skills hierarchy – think of chopping vegetables, packing boxes, dog walking, and helping elderly people go to the toilet. And this is why we need to collectively pause for thought before diving headlong in.

***In a nutshell:*** **In previous waves of technological innovation humans have shed redundant skills to climb higher and focus on what matters most in the new economy. This has also required a deeper investment in education and more people staying on until university, to extend the 'superpowers' of our pre-frontal cortex. But this is not necessarily an economic 'Iron law' baked into the fabric of the universe, for evermore (see Caplan, 2018). AI might unravel it quite soon.**

## 1.7 What are the shorter-term Implications for human skills development the AI Era?

As we outlined in the preceding section, in prior waves of technological innovation, humans have become more educated and moved further up the knowledge and skills ladders – often into evermore specialised domains.

In education and employment circles there has been an increasing and aligned emphasis on the so-called "21st Century Skills" such as critical thinking, creativity, collaborative problem solving, and communication. These all require facts and ideas, but also going beyond these to question, investigate, explore, and relate ideas that are claimed to matter more.

But as you might have gleaned from the preceding discussion on the capabilities of Large Language Models like ChatGPT, these systems are already incredibly good at these exact same things. They can:

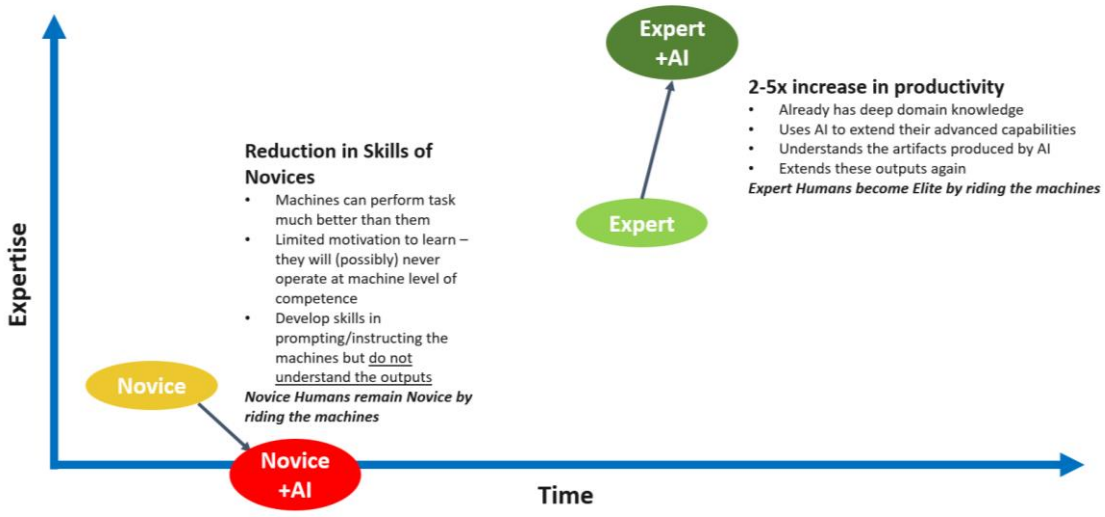
- Search for data

- Classify data and structure it into typologies
- Summarize what they have found and extract key findings
- Relate together seemingly unrelated ideas into sensible claims
- Generate high-quality long form text
- Re-write human text to improve it (or even write from scratch in the style of specific well-known humans)
- Make recommendations
- Assign probabilities to outcomes, even using Bayesian priors
- Write computer code, draw pictures, write music and film scripts
- Critique their own work and produce iterative improvements in seconds.

These seem very much like the skills that humans spend so long in the education system perfecting. Moreover, the AI models are constantly improving. We expect the next generation to be multi-modal i.e., you will be able to talk to it, send it videos, images, PowerPoint, long documents, music etc – and it, too, will be able to provide outputs in any of these formats seamlessly (see for example Microsoft Co-Pilot for an early glimpse of this). It might even have digital avatar that looks human, as you speak to it over 'video call', or it might remotely port into robotic bodies with distinct personas ("Sassy GPT" or "Serious GPT") – to interact with you in person.

We have no hard data on where this is likely to go in terms of the workplace. Our hunch, however, is that in the shorter-term it will make existing knowledge workers far more productive; but that it has the potential to stunt the capabilities of novices still at school or relatively early in their career. We present this in Figure 7 and unpack it further in the subsections below.

**Figure 7: Experts vs Novices in the World of AI**




***Expert Knowledge Workers***

These are the people with deep domain expertise in their employment field, likely acquired over many decades of on-the-job experience. Therefore, we think they may be more able to leverage AI to extend their existing capabilities – as a kind of ideas sounding board, cum-technical lead to write first drafts, cum-copyeditor to improve text, cum-quality assuror, and fact checker.

This would be almost akin to having a team of (digital) Harvard graduates at one's disposal 24/7 – except the AI is much faster than the human graduates and knows a lot more about a lot more

things! And they do not need feeding, will not come to work sleep deprived, are not rude or impatient, and do not feel superior to us. Unlike working with human colleagues, there is no drama, no hurt feelings, no office politics to contend with. You can ask the AI to do something, give it direct feedback on the draft, and then see it try again. Endlessly. Without breaks for coffee, lunch, family commitments, sleep, doctors' appointments, or the need for it to repay its student loans. All for a subscription fee of USD $20 per month (compared to a combined Harvard Undergrad and MBA cost of about US $300,000).

When experts work with these systems, it is likely that they will significantly extend their human capabilities. With their deep domain knowledge, they understand the outputs of the machine, can highlight the errors or areas for improvement and give fresh instruction. Much like the way an expert teacher might give feedback to a student. They can learn, too, from the interactions – gaining access to fresh ideas and perspectives that they had not previously considered.

Based on our findings from using these systems, our hunch is that in the short term, expert humans will become elite by "riding the machines". However, there may not be the same joy in prompting a machine and cutting and pasting/polishing the outputs as there is in crafting your own text, PowerPoints, pictures etc. from scratch. It fast becomes a world of human TV critics without any human screenwriters. But the Emmy-winning shows get written 5x faster, are much better, and eventually even get performed by digital avatars of long-dead actors.

***The Novices***

But what of the novices, the people currently in school that are preparing for the world of work? As a thought experiment, imagine you are 12 again but in today's world – with access to ChatGPT and a projected acceleration of AI unfolding at, say, 10x per annum. The risk is that the erosion of mental arithmetic, map reading, cursive skills etc. that we discussed in the previous section becomes the erosion of everything. Would we be motivated to learn things that machines can do in a fraction of a second, at near-zero cost? What would be the point of acquiring these skills?

Instead, we might focus on learning how to prompt the machines to generate useful outputs – just as many current postgraduate student's prompt statistics packages to run correlations and probability values – without knowing themselves how to calculate the scores manually and (often) without knowing the statistical theory and limitations of the different kinds of tests. But now imagine this applied to everything including writing love letters, doing your homework, and all doing your 'real' work. Sir Anthony Seldon calls this the risk of "infantilization" (Seldon & Abidoye, 2018).

It should be possible to put guard rails in place so that systems like ChatGPT do not give students 'the answer' but instead act like a Socratic teacher that helps them develop their understanding. But what's the point? If the AI can 'do everything' then every subject domain quickly falls into the existing trap of:

- **Latin:** "It's a dead language", = "So what's the point?"
- **Modern Foreign Languages:** "most people speak English" or "Google translate is already really good" = "so what's the point?"
- **The Knowledge** (i.e., that test of street-map knowledge required of London taxi drivers) = "but we can just use Google Maps or Waze and they know the fastest route, too"

"What's the point?" quickly applies to art, music, literacy, math, science and all the rest. Arguments about the intrinsic value of learning these traditional skills "for their own sake" may hold water for some time. But quickly, the skills become as quaint as Maypole dancing, fly fishing, archery, sword fighting, or crafting chainmail. They potentially become fringe activities pursued by hobbyists for fun – because they are no longer "mission critical" must-haves to survive and thrive in an increasingly automated world.

But if people no longer acquire these skills in literacy, math, and science – they may also struggle to prompt the machines or to understand the outputs of those prompts. Think about the sentence in Figure 8.

**Figure 8: Semantic Interpretation**

| Sentence: | Bush | Filed | a Suit | for Piracy |
|---|---|---|---|---|
| **Interpretation 1:** | The 41st or 43rd President of the United States | Submitted | A Lawsuit | For unauthorised distribution of copyrighted content |
| **Interpretation 2:** | A Rock Band | Put in a filing cabinet | | Ready for ship-based antics, conducted by men with parrots on the high-seas |
| **Interpretation 3:** | A British electronics company | Chiselled, shaved (or tailored?) | A type of formal clothing comprising jacket and trousers | |

It has multiple interpretations, and some are very silly. But we only know the correct one – because we have acquired and stored lots of useful background data in our long-term memory, that we can immediately recall. Without this learning (acquired through education), we become like the shop cashier when the till goes down – unable to add up the prices in our heads.

This means that if we want to stay useful in an era where machines can undertake most or even all cognitively demanding work, we will still need to learn lots of facts and information about a variety of domains and (effortfully) store these in our long-term memory, because we will need all this knowledge to prompt the machines effectively and to critique and improve the outputs.

**Bottom line:** you can't think, relate ideas, or be creative until you have some facts, knowledge, and ideas to think about, relate, and see differently.

But when we see the (ever increasing) gap between our own capabilities and those of the machines – it will take gargantuan reserves of willpower and grit to acquire this multi-disciplinary learning one step at a time. This brings us back to that "what's the point?" question and the risk and reality of mass downgrading human skills and abilities.

## Might we even forget how to read?

The global spread of literacy has been one of the greatest successes of the modern era. Around 15% of the global population was literate two hundred years ago, whereas today a similar percentage are illiterate (Our World in Data, 2023b)—one of the greatest achievements of the 20th century. These skills were spurred on by the invention of the Guttenberg press and the increasingly availability of books and newspapers that people wanted to be able to read. But the greatest impact was the spread of teachers (Hamilton & Hattie, 2022; Hattie & Hamilton, 2021). So, the incentives lined up and encouraged literacy to spread.

Fast forward to today and to the near-term advances expected in AI – "tomorrow" and the "day after". If you want to learn about something new you will soon be able to speak to your AI in words, rather than typing text and have a "proper" conversation. Your AI will be able to give verbal answers pitched to your current level of understanding. If you want more, it will be able to make bespoke videos in seconds – of the quality you might watch on the Discovery Channel. But in bite-size learning snacks. The AI will also be able to discuss these with you – Socratic style – to deepen and extend your learning. And, if you have smart glasses, you will be able to 'teleport' into a digital version of ancient Rome and have direct discussions with digital Caesar about his motivations from crossing the Rubicon and bringing down the Roman Republic. No need for pesky history books.

Before the invention of the Guttenberg press, most humans transmitted information and learning orally or visually – they watched someone in their tribe or village doing something. They tried to copy it and master the technique (Lancy, 2014). There were no 'how to' manuals – and this 'sorta' worked for over 200,000 years.

This is highly speculative, but **if we can have access to AI that can make Discovery Channel-like videos in seconds and even realistic immersive simulations – why would anyone bother to learn to read and write?** Yes, we can often process written words faster than their spoken counterparts, because the average speaking speed is 150 words per minute, but people can generally read 50% faster than that (Nation, 2009; Brysbaert, 2019). But would this matter much in a world where AI can do all the cognitively demanding tasks anyway?

Our provocation is that, within a few short decades, literacy skills may become as quaint as Latin and the Classics—things that we learn for bragging rights and the conferment of social status, but not in the least essential (or even useful) for day-to-day living.

Instead, oral communication may take on greater significance. The skills to work in groups, translate, undertake teamwork, and probe may well become more critical.

***In a nutshell:*** **it is not impossible that as AI capabilities grow, it could mass downgrade many human capabilities by reducing our incentives to learn. We might even forget how to read and write – as these skills would serve no useful purpose in day-to-day living.**

## Part Two: Four Scenarios for the Longer-Term Future of Work and Employment (and education)

Having now unpacked the capabilities of current AI systems and how these might quickly grow – and speculated on the possibility that this *might* result in a mass downgrading of humans, we want to present four (equally-speculative) scenarios for the future of work and employment. These are relevant to our world of education because one of the principal justifications for education funding is what the economists call 'human capital development' (Becker, 1994 [1964]). This is the idea that we should invest in school and university education to increase humans' critical thinking and creativity skills in an era where machines have automated farming, manufacturing, and most other types of 'blue collar' work. Although, as we have already alluded to in our preceding discussion, AI is now potentially snapping at the heels of the knowledge workers. So, in brief, here are the four scenarios:

1. **Future AI Developments are Banned** – with the technology being locked-in at its current levels.

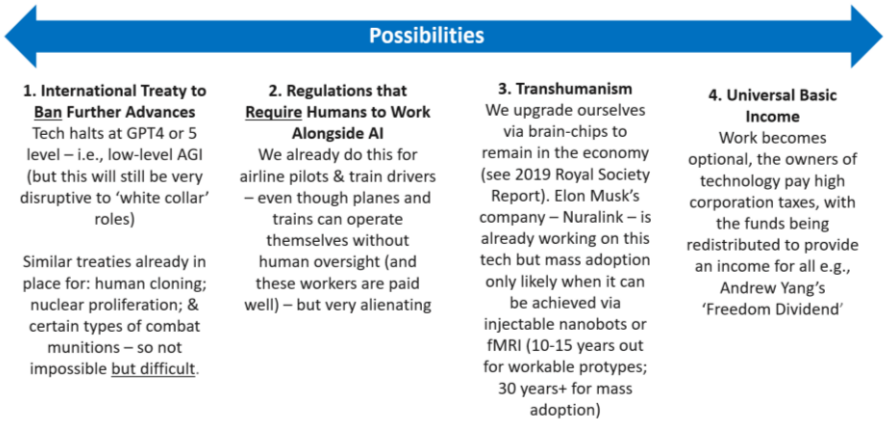
2. **Regulations that Require Humans and AI to work together (a.k.a. Fake Work)** – the technology continues to advance to a level where it can take over the majority of job roles, but with governments legislating to protect jobs and to require humans and machines to ride together.

3. **Transhumanism** – with humans upgrading themselves through "brain chips" and/or genetic engineering to stay relevant in a world of AGI.

4. **Universal Basic Income** – humans exiting the world of work and living a life of leisure, with a monthly stipend or 'freedom dividend' provided by the state.

To repeat, these scenarios are highly speculative. There are also other spaces on the board where humanity could land, it might vary by country/region, and these destinations are not mutually exclusive. We could also, initially, land on one of the four and then jump to a different tile. We summarize the four in Figure 9, below and then in the remainder of Part Two, unpack each in more detail. We do this so that readers can understand how we could inadvertently sleepwalk to one of these destinations – if we do not slow down, think, plan, and regulate RIGHT NOW. And then choose with great care.

**Figure 9: Four Scenarios of the Future of Work and Employment**



## Scenario 1: Future AI Developments are Banned

Let's imagine a world where policymakers in the US, the EU, and China become quickly and deeply concerned about the existential risks of AI—that the machines could quickly become all-knowing oracles that develop values and preferences different to our own. Although this scenario may seem totally far-fetched, many of those with considerable expertise in the area of AI are already deeply concerned about this possibility, including Max Tegmark, Nick Bostrom, Yuval Noah Harari, Elon Musk, Steve Wozniak, Yoshua Bengio, Stuart Russell, Michael Osborne, Daron Acemoglu, and the ultra-pessimistic Eliezer Yudkowsky (Future of Life Institute, 2023).

In this scenario, with governments fearing the short-term risks to electoral democracy, the longer-term risks of there (potentially) being no jobs for humans, and the even longer-term (and ultra-speculative) risks that the AI might decide to turn us all into paperclips (see Bostrom, 2014 for discussion of the paperclip maximiser thought experiment) – all major powers come together at a global conference and agree to ban any future development of the technology.

This seems highly fanciful but there have been several similar bans (or at least curtailments) in recent history:

- **Human cloning (2005)** – after the successful cloning of Dolly the Sheep and other experiments in the mid-1990s, the United Nations debated global regulation for around half a decade and ratified the *UN Declaration on Human Cloning* in 2005; with 70 countries then passing formal laws to ban the practice. Indeed, the only person known to have breached the regulations is imprisoned in China.

- **Treaty on the Non-Proliferation of Nuclear Weapons (1970):** this was designed to prevent the spread of nuclear weapons, promote disarmament, and facilitate the peaceful use of nuclear energy. It prohibits non-nuclear-weapon states from acquiring nuclear weapons and obligates nuclear-weapon states to work towards disarmament. Although not uniformly successful, it is thought to have contained the number of nuclear powers and reduced the number of deployable missiles each of those powers has at hand (Paul, 2020; Fuhrmann & Lupu, 2016).

Other lesser known but impactful global treaties include the Chemical Weapons Convention; the Biological Weapons Convention; Ottawa Treaty (Anti-Personnel Mine Ban Convention); Convention on Cluster Munitions; the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES); and the Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction (BTWC).

However, all these treaties took considerable time to negotiate, in many cases five or more years. The challenge we face with AI Large Language Models is that if, as some people estimate, their capabilities are growing 10x per annum (see Davidson, 2023 and associated literature on "AI Scaling Laws") then that's at least a 100,000-fold increase in capabilities during the period that nation states are debating the finer details of the treaty.

A second reason why we are sceptical that it would be permanently possible to halt AI advancement – you cannot uninvent it. It is much easier to progress AI research than many other areas. Building nuclear weapons, for example, requires uranium, which is very hard to get hold of, and building enrichment, manufacturing, and testing facilities cost tens of billions of dollars. Further, people notice – even when you do 'secret' underground tests. The seismic activity spreads for thousands of miles.

By contrast, OpenAI—the developer of ChatGPT—has fewer than 400 staff and according to some reports GPT-4 cost around USD $100m to develop. So, whilst this indeed is a large amount of money there are many private equity funds—and wealthy individuals—who could easily divert such an amount to funding AI development. And, unlike nuclear testing, no one would know.

Regulation is not totally impossible, however. The global network of companies that design, manufacture, and supply computer chips is remarkably small—just five organizations (Miller, 2022). This means that in the short-term, decisive action by a small number of governments might be enough to control both the supply of chips and access to compute that provides the 'brainpower' to Large Language Models – giving additional time to negotiate a global treaty.

However, our hunch is that, without immediate upward pressure from citizens, it might take even this small number of governments several years of collective coordination to act to stop further training runs and advancements in LLM capabilities. By that time, we may already be at ChatGPT-6 or 7 and decentralized opensource models, with capabilities approaching AGI – and with profound implications of the future of employment—proliferating.

However, even if a global treaty is successfully passed – or some other form of curtailment, such as constraining the global supply of computer chips is agreed – it will likely be more successful at stopping the public proliferation of advanced AI. The continued private development of such systems might well continue, given the relatively low entry costs. It is also possible that LLMs might become more computationally efficient and no-longer require state-of-the-art chips.

Nevertheless, it would still give us all more time to think, debate, and decide which of the other three scenarios we might collectively prefer.

---

Of course, there might be some advantages to ever increasing automation. In many advanced economies, birth rates are plummeting, and the tax coffers generated from the younger working generations will not likely be enough to fund the older retirees (United Nations, 2019). For some countries, ever-increasing numbers of 'digital people' might be preferable to opening their borders to large numbers of physical people with different beliefs, values, and ways of life.

It is also probable that, with the help of ever-greater AI, we will be able to solve some of the world's most pressing problems such as climate change, disease, water shortages, and disparities in international development with much greater speed. Many, such as venture capitalist Marc Andreessen (2023), argue that rather than slow down AI advances – we should instead speed up – pushing ever harder on the accelerator. We think that, eventually, there may indeed be a case for speeding up. But we should only accede to this *after* we have put sufficient regulatory guardrails in place.

## *Implications for Education*

If, however, we can keep the AI genie partially in the bottle – stopping ever more powerful models from being released into the wild – we think that the implications *could* be quite positive for education. For example, if Large Language Models get stunted after "GPT-4.5", we may have the best of all worlds. Humans remain firmly in the driving seat and that continued agency gives

strong incentives to stay invested in getting an education. We still need those advanced cognitive skills to make better decisions!

We might also see stronger guardrails being placed around the AI, enabling it to be an accelerator of human educational excellence. These might include some or all the following:

1. **Lockouts that prevent students from using the systems for cheating** – think of this as a sophisticated version of parental control systems on video streaming platforms that restrict what children can watch. This would be even more advanced, tracking what children are looking at on their devices and literally preventing them from prompting the AI to give them answers to their homework. Instead, every time they try to cheat, they get Socratic responses and suggestions for how they could (and should) do the work themselves. Given that even AI finds it difficult to tell whether text has been written by another AI, we will have more luck in simply preventing students from getting the machines to write their homework through lockouts than in trying to check afterwards. (Although we also note that AI companies are also exploring use of 'digital signatures' so that plagiarism detection software can identify AI-generated text with greater accuracy)

2. **Fact-Checking Apps** that enable us all to overcome the growing number of deep-fake videos and dubious op-eds that are contributing to increased political polarization. Enhanced education can support people to verify the truth behind what they see and hear, but we think there is also a role for 'good' AI to help defeat 'bad' AI. This 'good' AI could pre-scan every article that we read, colour-coding the text – depending on whether it was an undisputed fact, a widely accepted fact, an acceptable interpretation of facts, a spurious interpretation, or a downright lie. Early versions of this might look like the spelling and grammar checker wavy lines in Microsoft Word – with different colours corresponding to the degree to which we can trust and accept what is written. They might also provide us with citations and links to justify the fact checking colour rating that has been applied.

   Later versions might analyse video and display a coloured dot in the corner of the screen that changes colour moment by moment – depending on the 'truth level' in the words that are uttered by the people on screen. And for both the text and video checker, an overall 'Accuracy Score' for those that do not have the time or inclination to explore the truth level, line by line. Yet another way of doing this would be simply to ask the AI to 'remove all bias' from the source text – re-writing it to filter out dubious opinions.

3. **Teacher workload reduction technology**. There are currently very high attrition rates amongst teachers in many advanced economies. One of the commonly reported reasons for this is excessive teacher workload. Indeed, we have recently published on book on the science of de-implementation for education, to help educators get much needed work-life balance, without harming student learning (Hamilton et al, 2024). But we can see significant potential for AI to help too. Imagine an integrated App that:

   a. **Produces lesson plans, teaching and learning resources and student homework tasks** – personalizing much of this to the needs of individual students

   b. **Marks homework and essays and provides feedback** – perhaps also monitoring and tutoring students, step-by-step in real-time, as they undertake their assignments and giving them prompts and feedback

c. **Acts as a personal assistant cum gate keeper with parents**, handling parental queries and giving them an endless stream of real-time data, should they seek it

d. **Supports teachers to identify their areas for personal development**, helping them to establish implementation intentions, then monitoring and 'nagging' them, and helping them to evaluate the impact

e. **(optionally) enables teachers to wear Augmented Reality Glasses** – with a teleprompter containing their lesson materials and a feed showing the biometric data for each student, giving inferences on the respective levels of engagement and learning; and in the moment feedback and suggested next steps

f. **Helps school/university leadership teams with their improvement planning** – by analysing data to recommend priorities, undertaking root cause analysis, proposing, stress testing, supporting and then evaluating initiatives.

Such technology could substantially enhance the effectiveness of educators, whilst also giving them much-needed work-life balance. It could, of course, also significantly erode professional expertise – with teachers and lecturers expertly reading their lines but not knowing why particular approaches work.

4. **A Personal Digital Tutor for All Children.** This could either be aligned with the teacher support App that we have described above, or be a standalone system, and it would be able to act as:

a. **A Sophisticated Intelligent Tutoring System** that constantly monitors each learner's past and current level of progress, creates bespoke learning content/activities, assesses progress, provides in-the-moment feedback and encouragement, and then makes optimal choices about next lessons. This would augment the expertise of human teachers in advanced economies and might also democratize access to high quality personalized education for children in the poorest countries, where education provision is often still woefully inadequate.

b. **A Digital Friend and Coach** that students could rely on as a source of advice, guidance and as a critical friend. This could also help them develop their hobbies and interests and connect them to new friends in the real world, booking 'playdates'; and – if they are lost for words at the first meeting—even delivering scripted lines in their AI glasses. These digital coaches might also encourage healthy behaviour like 'eating your greens' and influencing them not to take up vaping or to skip classes.

It will be important, however, that these systems prioritize student agency, help learners to come to their own decisions, help them to find and interrogate their own information, and basically help them to learn how to learn. In particular, students will still need teachers to help them make wise choices about the next step in their learning (as novices do not know what they do not know).

## Scenario 2: Regulations that Require Humans to Work Alongside AI (a.k.a. Fake Work)

This scenario assumes that governments fail to put the AI genie back in the bottle—that they either take too long to coordinate and regulate, or that that regulation simply slows down the inevitable and that the dam eventually breaks, with intelligent systems proliferating.

As AI systems become ever-more powerful, they won't just be able to do just bits and pieces of existing jobs. It's very likely that they would eventually be able to take on the whole role: the analysis; the planning; the execution; and even the interpersonal interactions with humans. We illustrate this transition from time-saving devices to knowledge partners, and finally to all-knowing Oracles in Figure 10, below:

**Figure 10: The Evolution of Technology**

| Level 1: Time Saving Devices | Level 2: Knowledge Partners | Level 3: All-knowing Oracles |
|---|---|---|
| Washing Machines; Combine Harvesters; Word Processors; Google Maps; Calculators etc. | Current generation Large Language Models e.g., GPT-4; Bard; LaMDA etc. | Next generation Large Language Models that achieve general or superintelligence |
| ***Technology does the dirty work so we can do the knowledge work*** | ***Almost our equals, the technology amplifies our mental capabilities and helps us to fly*** | ***We struggle to understand what they are talking about; and need them to dumb down and explain everything in 'simple' terms that a 60- year-old human with a PhD could understand*** |

These machines may still lack "consciousness" and have no real "emotions" – but might not matter that much. This likely won't stop them from being able to fake it and for their interactions with us being indistinguishable from our interactions with one another. And in some contexts, it might even be better for them not to fake it. In our world of education, the current experience of students being taught by robots show that students often prefer to be taught by a non-emotional and non-judgemental machine (Hattie, 2023). The AI 'teacher' does not know they are the naughty child, the one on the spectrum, it does not label, and does not roll its eyes and get frustrated when asked it to explain the same concept a thousand times. It also does not smirk when you get it wrong.

Indeed, some (see for example, Seldon & Abidoye, 2018) have speculated that teachers could perhaps be entirely replaced with AI before the end of the current decade. These systems would have intimate knowledge of each student, deep content knowledge, and advanced capabilities to weave appropriate pedagogies and evaluate their impact on student learning. They could make learning visible, including using biometrics to have a good sense of what has 'gone in', how deeply, and when and how to engage in spaced repetition to deepen understanding and application yet further.

But, of course, the implications are not just for education but for efficiency and automation in the wider economy – so it's a two-way street. Particularly because one of the principal justifications for

funding education is to prepare children for the world of work in the first place. Most readers of this paper likely live in some sort of market economy where competition drives efficiency.

Recent labour-market analyses by Goldman Sachs (2023); McKinsey (2023); and Open AI (Eloundou, et al, 2023) – suggest that up to 300 million jobs are already *partially* automatable by existing AI, with most of these being higher-paid knowledge work roles. But these reports also make predictions about new jobs. Their assumption is that even though many existing roles might be automated, AI will catalyse innovation that will result in many new jobs, products, and services that we can't yet imagine.

We agree with *some* of the reasoning but not all. Yes, it is likely that there will likely be many more products and services that emerge courtesy of the AI, that we could not even have dreamed of. Yes, too, in previous waves of technological disruption the jobs created have indeed outnumbered the jobs destroyed. These jobs have often been much higher paid resulting in our collectively improved standard of living. But we do not necessarily agree that new roles for humans will *always* emerge from this process.

If as a result of these AI advances you have access to "digital people" that are as good – and possibly better – than organic people, that can function 24/7, without holidays, or even payment – and you were on the board of directors of a publicly listed company (or even a school board), what would you do? It's only natural that you might at least *consider* replacing the organic people with the cheaper and less error-prone digital people. Particularly if you could spin-up the thinking power of thousands of "digital people", each with the 'horsepower' of, say, 10,000x Einstein, at a fee of a few dollars per digital person, per month. These digital people would not require inductions, training, health and safety audit, wellbeing support, or to go to teambuilding lunches. Some of them would also be Super-Eddisons– who manage the more academic Einsteins to get things done; no need for human supervisors either!

Of course, this transition might not happen that quickly. First, we need the technology to get to this level (which, according to the bullish is only 2 years out vs the more bearish estimates of 2030-50). Second, we need widespread adoption – including sufficient infrastructure for scaling. Yet, one of the things about us humans is that most of us empathise with one another; senior leaders in organisations generally don't typically enjoy making people redundant. If times are good and they can carry the surplus (human) baggage, we think they will likely seek to do so: having humans and machines riding together. Instead, when economic downturns occur – and organizations enter survival mode – they are more likely to consider radical change. (Although we also prefer receiving services from people rather than machines – another variable that may slow things down).

However, we might see two dynamics at play:

> (1) a gradual 'boiling of the frog', with human workers (including teachers) all having an AI digital assistant that supports them – but which could probably do ALL of the human's role (it's learning the ropes by shadowing and supporting you in your role);
>
> (2) an economic downturn resulting in the layoffs of an ever-increasing number of human workers – particularly highly-educated knowledge workers (see Eloundou et al, 2023 – for a recent labour market analysis of the impacts of LLMs on human employment).

Think here of large global enterprises that currently employ hundreds of thousands of humans – perhaps gradually culling back to a workforce of less than 500. With many of these being in insight roles – helping the machines to understand how humans think, what products they might like, and

what messages might appeal. Those humans might end up being little more than an (expensive) in-house focus group – with even the organizational leadership roles largely being undertaken by AI.

If this were to come to pass, the implications would (at least initially) be significant. We humans spend a considerable portion of our days working and derive a large part of our identity, social status, and social bonds from our jobs, job titles, and the prestige of our respective organisations. And we also invest considerable time and money in going to school and college to get the qualifications that permit access to all this.

If work came to an end, it would potentially upend the fabric of society, as previously busy and productive people instead sat at home watching daytime TV or playing computer games; and struggled to find another source of meaning a purpose in their (relatively) stress-free lives (see Scenario 4 for the potential reasons for that lack of stress).

Governments are likely to be very concerned about the impact of such a radical transition. They might choose to legislate to slow it down or to stop it completely, which brings us to the heart of this scenario. There are several different ways they could undertake this slowdown, including:

- **Fines** e.g., for every human whose job is eliminated by technology, organizations would have to pay a penalty (perhaps as much as 10x their annual wages).

- **Sliding Scales** e.g., a formula that links the required number of humans per company to the annual revenue turnover and surplus. This might also include tax breaks and incentives where the organization exceeds the legislated minimum human headcounts

- **Linking human headcounts to the amount of digital computation undertaken** e.g., using metrics that extrapolate the quantity of AI computation used by each company and converting this back into 'human working days', to calculate how many humans would be required to achieve the exact same outcomes. And requiring companies to have this same level of human headcount, 'working' alongside the machines.

Each of these regulatory regimes can be thought of as a forced job-creation scheme, almost akin to John Maynard Keynes' example of paying people to dig up roads and then fill them back in – *ad infimum*.

This may seem alien and odd but there are precedents to this in the modern economy already. Trains no longer really need drivers. When the Victoria Line on the London Underground opened in 1968, the trains could drive themselves. But in 2023 we still have a relatively well-paid person sitting at the front. Similar arguments have been made for commercial air travel. The autopilot is usually engaged for 90% of the flight and aircraft already have the capability for totally automated landings. It's (currently) just take-off that requires a pilot but there's no reason they couldn't be sitting in the control tower at the airport, remotely piloting the plane for those few minutes. And there are already AI-driven planes, although these are not (yet) in commercial use.

In his 2019 book *Bullshit Jobs: A Theory*, David Graeber contends that more than 50% of existing jobs are utterly pointless. He divides these roles into the categories of flunkies, goons, duct tapers, box tickers, and taskmasters – suggesting that these inefficient roles proliferate because of a "managerial feudalism". That is, senior leaders derive status and power from the size of the departments they lead, and much of the 'work' of this retinue actually adds little to the effectiveness of the organization.

Extending from the examples above, one possibility is that as AI can undertake all cognitively demanding work to a higher standard than humans, and as companies increasingly seek to reduce the number of employees, governments simply ban this from happening. That they make it a requirement that the corporations maintain a full complement of human workers – including teachers – and pay them a salary to work alongside the machines.

We can, however, see some frictions or challenges with this:

1. **The machines would clearly "think" a lot faster than humans and would not need to attend endless meetings and generate large amounts of correspondence to come to a decision**. In fact, they may well have determined the best course of action in a matter of milliseconds. If humans require several days and a team building event to make a decision and ultimately end up agreeing with AI analysis anyway, it is hard to see what contributions the humans are making.

2. **The degree to which the human employees might quickly become infantilized.** We could imagine a world where the technological advances discovered by the AI occur at such pace that the humans can hardly keep up or understand. Instead, everyone wears augmented reality glasses (including teachers) and reads the lines produced by their AI personal assistant when attending meetings or classes (i.e., act out). Literally no one at the meeting (or in the class) has the faintest idea what they are talking about – but each actor delivers their lines, pretends to understand what the others are saying and then declares the meeting closed and the decision made. Of course, this assumes that people can still read. If they cannot, then the meetings will run slightly slower, as people wait for their AI to whisper talking points into their earpiece – for them then to repeat. Or it might simply adopt their voice to speak "on their behalf".

3. **A deep sense of alienation.** A case in point: when we wrote this paper, we used ChatGPT-4 as a digital assistant. The outputs were of a good enough standard that we could have almost cut and paste them into the paper, and it certainly would have been more efficient. That said, we like to believe that our careful writing and editing has produced a better text. However, we can certainly imagine that the outputs of LLMs will improve to the point where all our attempts to make the text our own make it worse.

4. **Some companies might deliberately automate as much as possible, as quickly as possible, for competitive advantage.** To flesh this out, one of the key success criteria for businesses is to get high-quality products and services into the market faster than their competitors. As James Gleick (1999) pointed out over 20 years ago, in many industries, projects that run 50% over budget but finish on time are more profitable than those that keep to budget but arrive late. But human deliberation (meetings, emails, office politics etc.) slows this all down. So, some companies might seek to bypass this by having an entirely AI-workforce to get from strategy, product development, testing, and manufacturing as quickly as possible. Maybe even in minutes. This would create an "arms race" in which other companies would be forced either to follow suit, or risk being left behind. Some education systems might opt for this path too – especially given that education becomes much lower stakes in a world without jobs.

5. **Whether the companies might end up treating the requirement to hire humans as a sort of closet-welfare payment; and preferring that their "employees" do not actually come into the office or do any work.** In recent history there were precedents to this in some of the Gulf Cooperation Community Countries – where it was common for international firms to have to hire a quota of locals alongside the "expats". Sometimes these local employees were

encouraged not to come into work and the system was arguably an "arms-length welfare benefits and dignity machine", administered indirectly by the companies.

There is also a profound difference between David Graeber's contemporary example of 'Bullshit jobs' and AI-induced 'Fake Work'. In the contemporary version it's not totally clear that the work is pointless and there will always be some aspect of even the most tenuous jobs that will enable the postholders to say: "I made a difference today". But in a world where the AI can do everything better than us – everyday becomes "I slowed things down today: I got in the way".

### *Implications for Education*

Theoretically this scenario requires zero education – beyond learning the civic niceties. The human employees are merely organic simulacrums – playing the parts and saying the lines prepared for them by the machines, in a (fake) job-creation scheme. Much like the ancient Athenians who selected citizens for political office by drawing lots – we could all take our turn acting out the role of "managing director", "chief secretary to the treasury", "teacher" and "school leader". If we all craved status, we could all be co-managing directors – in a wider symphony, speaking the notes crafted by the machines. This would mean that there would be little purpose in high-stakes student assessment. There is no need for a sorting hat to divide the school leavers into janitors, professionals, or senior leader material. We could each be Prime Minister or President for a day and get to enjoy the motorcade – as long as we don't deviate from the (faultless) AI script. And if we keep schools – better for teachers to follow the script, too.

## Scenario 3: Transhumanism

One of the clear challenges with the previous "humans riding the machines" scenario, is that as the machines become ever more intelligent, we play second fiddle to their brilliance. We are never the architects of breakthroughs, instead we are just 'in the room' when it happens and only have an extremely hazy understanding of what is going on. And we are just getting in the way and slowing everything down.

Our strong suspicion is that, for many, this will not be enough. That 'being there' will feel hollow, almost akin to being a non-player character in a computer game; relegated to being the sidekick rather than Player One, driving the action.

It may therefore be tempting to 'upgrade' ourselves so that we can hold our own with the machines and partner with them as equals. There are (at least) two paths to this:

(1) genetic modification; and (2) implants.

The *genetic modification* path would be about amping our natural organic capabilities – either by assortative mating (i.e., smart people seeking to make babies with other smart people to increase the chance of even smarter offspring) or by editing the 2,000+ genes that influence IQ e.g., with the existing CRISPR gene editing suite (Hamilton & Hattie, 2022; 2023; Bostrom, 2014).

We are sceptical that genetic modification would take us far enough, however. Assuming humanity could get past the (very reasonable) taboos on eugenics, assortative mating would likely take hundreds of years to generate sizable increases in human intelligence. And whilst gene editing would

work much faster, our hunch is that it would quickly hit the limits of our biological substrates. Yes, we might be able to think much faster and have a larger working memory and all with much lower power consumption that computer servers. But unlike silicone devices, our organic neurons would not operate at the speed of light, would still need sleep to remove the daily toxin buildup, and we would only be able to 'consciously' work on one thing at a time while our AI friends can think about millions of things at once. We would still be very much the slow coaches.

Therefore, we think a second and more likely path would be for people to have digital implants as envisaged by Arthur C. Clark (1998) —to become hybrids that are part-human and part-machine.

Back in 2016, Elon Musk founded Nuralink—a 'brain-chip' implant company—for precisely this purpose (Fiani et al, 2021; Agnihotri & Bhattacharya, 2023). Musk's hunch was the one-day AI would greatly exceed the capabilities of organic humans and that some of us might welcome the option of being able to upgrade ourselves.

Nuralink and other similar technologies are still very much in their infancy. They work by surgically implanting a web of bio-compatible probes into the brain that can be used for two-way transfers of 'data' but of course we can expect major advances in the coming decades.

Our hunch is that as long as having a brain-chip requires surgery—which is the case with the current generation of experimental technology—most people (us included) would not wish to be upgraded in this way. However, many scientists and futurists predict that with advances in nanotechnology, cognitive enhancements could be produced by simply drinking a glass of water infused with millions of nano-bots or *neural dust* (Shanahan, 2018; and Royal Society, 2019). Each of these bots might be the no more than 200 nanometres (0.0002 millimetres) across and once in the bloodstream this armada of tiny devices would then end up in your brain and attach themselves at key neuronal junctions and act like a 'WiFi network'.

This technology is still just theoretical, and we have no way (yet) of knowing what it would feel like to have this kind of upgrade or about how it would change us. But some considerations are as follows:

1. **Our brains could theoretically become part of a wider networked WiFi network.** This could let us download new information and skills from the cloud in seconds. To us, it would seem like we always knew the stuff in question – whether it be astrophysics, Mandarin, or the complete works of Shakespeare, because the cloud would literally become an extension of our long-term memory.

2. **As long as our WiFi connection is on and strong, we would each have mental processing powers hundreds or thousands of times that of Albert Einstein.** Theoretically we may be able to 'consciously' think about several things at the same time – much in the same way that ChatGPT is able to process millions of user requests simultaneously (this would increase the human demand for fresh content, because we can now watch the entire Netflix library at once). But should the WiFi be cut, we would be back to our Lo-Fi selves

3. **If we can Bluetooth to the cloud, there would also be no reason why we couldn't Bluetooth to each other.** Instead, we might no longer need to speak – opting for more efficient data transfers to discuss and combine ideas, with all this occurring digitally in tenths of a second. The end of long-winded meetings, chitchat, and gossip—all replaced with a quick burst of binary code.

4. **But there would be major risks to brain-hacking.** Already, many of us have concerns over the level of privacy controls on our social media feeds and fears that our smartphones are listening to us, when the thing we were talking about 5 minutes ago suddenly appears in our advert feed. With brain-chips, however, the risks become far more Orwellian. We face the very real risk that App developers and state-actors can monitor our thoughts directly; and even implant ideas or move our limbs remotely. There would need to be cast-iron regulation and foolproof anti-mind-hacking software to guard against this.

5. **There would also be major philosophical questions about whether we were still human and what we had lost and gained in upgrading ourselves.** Much like Thomas Nagel's (1974) essay on *What is it Like to Be a Bat?* – we could not know what it is like to be an upgraded human until after we had had the upgrade and the process might not be easily reversible, if we do not like what we have become.

6. **But an arms race might push us all to get the upgrade.** Those that have 'the chip' become – overnight – more cerebral than the combined might of every Nobel Prize winner that has ever lived, while those without 'the chip' would become prospectless, relegated to the animal class.

7. **And we still wouldn't be as good as the machines.** Our protein-based brains cannot process information at the speed of light and require daily rest for removal of neurotoxins. So, our chips might merely upgrade us from "special needs" to "average" level. One consolation might be that human brains are (currently) much more energy efficient than supercomputers and that our bodies can automatically repair themselves, which machines cannot!

We are aware that all this talk of brain-chips may seem like impossible science fiction, but it is important to realize that there is a strong consensus in the fields of human cognition and artificial intelligence that all this is eminently possible and little more than a 'technical problem' that could be overcome with the help of ever more capable AI systems (Shanahan, 2018; Royal Society, 2019).

Already researchers have established one-way links to human brains with humans sitting inside MRI machines whilst silently reading a book (Tang et al, 2023). When advanced AI-systems monitor those brain activities and compare them with the passages of the book we are reading – they can (and do) learn which parts of our brain are active as we think different words; and they can report back the 'gist of it' with reasonable accuracy. If we can then miniaturize these MRI probes into daily wearables – we can write essays or do work with our minds, without any need to tap the keys. This is only the start, and we think that two-way transfer will (eventually) come.

## *Implications for Education*

The implications of all this for education are both stark and very obvious. Why would anyone need to go to school or university when they could download to their brains anything they needed to know from an Appstore in mere seconds? Whether that be learning how to drive a car manually; speak Arabic; or interact with regular (i.e., non-upgraded) humans with compassion and empathy.

## Scenario 4: Universal Basic Income

Our final scenario assumes that the genie is not put back in the bottle and that humans reject both the notion of Fake Work and the risk of upgrading their brains and transcending their humanity. Instead, we become completely de-coupled from economic activity. The machines do all the work, all the thinking, all the deciding and all the innovating. And humans get to enjoy the fruits of this silicon-mechanical labour.

One pathway to this endpoint might be that as corporations exploit the efficiencies that artificial intelligence allows, they shed ever increasing numbers of jobs. They do this because human involvement in decision-making slows down the rate of progress because the machines have to spend too much of their time explaining, seeking agreement, and getting signoff from humans. And, of course, because the machines require no salary, no off days, and are not unionised (exactly as we outlined in Scenario Two, although in that earlier scenario the governments were attempting to play whack-a-mole, to regulate against this and keep people in jobs).

However, one of the key reasons that our existing economic system is successful is that, through employment, people earn wages that they can spend on the products and services developed by others. If, on the other hand, all – or most – of the human workers have been laid off, they no longer have the resources to meet their basic needs, let alone satisfy their taste for life's luxuries. And without customers, all the corporations then collapse—no one can buy their products.

One way around this is to tax the corporations at very high tax rates and redistribute the funds to citizens as a kind of Universal Basic Income (UBI). There have been many proponents of UBI over the years, including Philippe Van Parijis (1995; 2017), Guy Standing (2017), Martin Fford (2015), and Andrew Yang (2018). Several different ways of funding UBI have been proposed including corporation tax, profit sharing sovereign wealth funds, data dividends (where people are paid for the data held about them in cyberspace), and a job-replacement tax levied on companies based on how many human roles they have eliminated and how many they create.

There are even more radical versions of this scenario, such as Aaron Bastani's (2022) *Fully Automated Luxury Communism*, which speculates that AI competition may eventually lead to large monopolies that governments could then put into common ownership. With the utopian suggestion that all goods and services could then be totally free to citizens, because the production cost would reach near zero – with all the thinking done by AI and all the hard work by robots. There can be little doubt that AI systems would be better at centralized coordination and planning than the communist regimes of the past.

Many of the UBI scenarios could result in all humans leading pampered lives. But in a context where the machines become more and more capable – we might feel like ants conversing with an all-knowing Oracle. Yuval Noah Harari (2016) speculates that many humans might even worship the AI like a 'god' – albeit one that attentively listens and responds to our every thought with useful and actionable suggestions.

This economic de-coupling scenario might even result in many government functions being increasingly transferred to AI, as we come to realize that our sleep deprived human leaders often make poor decisions.

With a UBI, we would have the flexibility to decide what to do with our days, whether that be gardening, socialising, creating music in choirs and musical groups, participating in historical re-enactments, cultural conservation, or porting into the metaverse to temporarily experience other

'realities' and play games. In other words, it would be quite a lot like being in the permanent state that we currently call "retirement".

There might also still be social status in owning luxury goods that are hand-produced by other humans – even if these are actually inferior to AI produced goods. After all, many people like to buy expensive mechanical watches that are much less accurate than cheaper quartz watches, so we might also see a proportion of people opting to become artists, makers, and tinkerers to overcome boredom and to give their lives meaning; and then selling their (high-status) produce to others. This would generate a corresponding need for vocational education for them to learn their tradecraft.

### *The Implications for Education*

The implications for education with this scenario are less stark than with transhumanism. We might still go to school to socialize, play games, learn basic skills, and to challenge ourselves against other humans – much like the current crop of human chess players. But human capital development would no longer be a priority. We would no longer be preparing for high stakes exams to gain entry to prestigious employment opportunities. Although for some, there might be a focus on learning vocational skills to develop and sell human-made products, such as hand-turned wooden bowls.

Our model of schooling might then look more like the Sudbury Valley paradigm (Gray, 2013), where children decide themselves how to fill their days, what to learn and what to tinker with. There is also the option of accessing a skilled adult (or an AI) for support, when they feel they need it. It might also involve the requirement to learn the ever more complex formal rites and rituals that might fill our days and offer meaning and fellowship.

Indeed, when the Aztecs first invented universal education in the 14th Century, its purpose was largely to learn the key public rituals and the memorize songs and poetry that told them of their past, and their relationship with their ancestors and their gods (Reagan, 1994). Education in an era of human de-coupling from economic activity might look very similar.

Humans would focus on human things like love, compassion, friendship, social support, and play. We would probably become more interested in art, music, dance, drama, and sport, learning how to do these things to an elite (human-level) standard. After all, even if machines could do these things to a higher standard than us, we wouldn't be terribly interested in watching them.

## PART THREE: Where do we collectively go from here?

## 13 Recommendations

By this point, you may be deeply sceptical or even perturbed by the scenarios we have discussed. You may feel that the mass deskilling and de-education of humanity seems like something that could never happen. You may say, "schooling has been through these existential crises before and survived" although in this context it is worth noting that the current model of schooling is but 150 years old.

However, we feel we have actually rather pulled our punches by focusing on the implications of human learning, human agency, human employment, and human fulfilment. There are many other reasons that we should also be worried (or at least on high alert) about the latest advances in AI. This includes the potential for:

- Electoral interference (already being discussed by the US Senate Judiciary Committee)
- Spreading fake news
- Proliferation of online scams
- Radicalizing people's opinions, with even greater effect
- Rogue states using AI for Orwellian 'thought policing' and human enslavement
- Autonomous weapons
- Increased concentration of power and wealth
- Super-intelligent systems developing goals of their own and coming into conflict with humans.

But of course, there could also be major benefits from these systems if development proceeds carefully, including:

- Addressing climate change, for example by the development of low-cost carbon capture technology and nuclear fusion
- Enabling everyone to have a personalized world-class education
- Extending human lifespans
- Finding cures for all known diseases
- Reducing military conflict by "gaming out" wars in the digital realm so that the AI is able to tell us with high accuracy who would win before troops are committed and at what cost the victory would come
- Accelerated exploration and colonization of other planets, so that we don't have 'all our eggs in one basket'.

Given all the above, we think that banning further developments in artificial intelligence would be unwise – but we feel strongly that need to be very carefully regulated, so that we can slow things down a little. To then, carefully and collectively decide which vision of the future we like the most; rather than accidently sleepwalk off a cliff from which there is no coming back.

Our 13 recommendations are as follows:

1. We should **work on the assumption that we may be only two years away from Artificial General Intelligence (AGI)** that is capable of undertaking all complex human tasks to a higher standard than us and at a fraction of the cost. Even though AGI might still take several decades, the incremental annual improvements are still likely to be both transformative and discombobulating and this requires our collective attention *now*.

2. Given these potentially short timelines, we need to establish quickly a **global regulatory framework** including an international coordinating body and country-level regulators.

3. That AI companies should go through **an organizational licensing process** before being permitted to develop and release systems 'into the wild' – much like the business/product licensing required of pharmaceutical, gun, car, and even food manufacturers. This licensing regime would allow approved AI companies to build and tinker with systems in their 'labs' and to undertake small-scale testing – under the supervision of the licensing agency.

4. **End-user applications should go through additional risk-based approvals** before members of the public are given access, much in the same way as pharmaceutical companies have to get drugs licensed and food manufacturers demonstrate their products are safe. These trials and tests should be proportionate with the risk or harm, with applications involving children, vulnerable or marginalized people subject to much stricter investigation.

5. **Students (particularly children) should not have unfettered access to these systems before risk-based assessments**/trials have been completed.

6. **Systems used by students should always have guardrails in place that enable parents and educational institutions to audit how and where children are using AI in their learning.** At the very least, we should expect that children would require permission from parents and schools prior to being able to access AI systems.

7. New laws to make it **illegal for AI systems to impersonate humans** or for them to interact with humans without disclosing that they are an AI.

8. **Measures to mitigate bias and discrimination in AI systems**. These could include guidelines for diverse and representative data collection and fairness audits during the LLM development and training process.

9. Stringent **regulations around data privacy and consent**, especially considering the vast amounts of data used by AI systems. The regulations should define who can access data, under what circumstances, and how it can be used.

10. **Require AI systems to provide explanations for their decisions** wherever possible, particularly for high-stakes applications like student placement, healthcare, credit scoring, or law enforcement. This would improve trust and allow for better scrutiny and accountability.

11. As many countries are now doing with the Internet systems, **make the distributor responsible for removing untruths, malicious accusations, and libel claims** – in a very short time of being notified.

12. **Establish evaluation systems to continuously monitor and assess AI applications' safety, performance, and impact**. The results should be used to update and refine regulations accordingly and could also be used by developers to improve the quality and usefulness of their applications – including for children's learning.

13. **Implement proportionate penalties for any breach of the AI regulations**. The focus could be creating a culture of responsibility and accountability within the AI industry and end-users.

**We are aware that there will be strong differences of opinion about many of these recommendations** and for that reason, we suggest that you treat them as stimulus to further debate, rather than as a final set of cast-iron proposals. But we need to have that debate FAST and then enact pragmatic measures that give us breathing room to decide what kind of future we want for humanity.

These are undoubtedly complex issues that require much dialogue, involving policymakers, AI developers, and society as a whole. There are bound to be wildly diverging assessments of the risks and opportunities. That is why we think it is essential to adopt a multistakeholder approach in reviewing the AI opportunities and risks, creating robust and practical and proportionate regulations, especially given the risk that those regulations might inadvertently create global monopolies amongst the first AI movers.

But we need to get moving now!

# Bibliography

Agnihotri, A., & Bhattacharya, S. (2023). *Neuralink: Invasive Neurotechnology for Human Welfare*. SAGE Publications: SAGE Business Cases Originals.

Andreessen, M. (2023). *Why AI Will Save the World*. Andreessen Horowitz, Why AI Will Save the World | Andreessen Horowitz (a16z.com)

Armstrong, S., Bostrom, N., & Shulman, C. (2016). Racing to the precipice: a model of artificial intelligence development. *AI & society*, *31*, 201-206.

Asimov, I. (1950). Run Around. I, Robot (The Isaac Asimov Collection ed.). Doubleday, New York.

Barrett, L. F. (2017). How Emotions are Made: The Secret Life of the Brain. Houghton Mifflin Harcourt.

Bastani, Aaron. *Fully automated luxury communism*. Verso Books, 2019.

Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

Bowman, Samuel R. (2023). "Eight Things to Know about Large Language Models". arXiv:2304.00612

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language Models are Few-Shot Learners. arXiv preprint arXiv:2005.14165. Retrieved from https://arxiv.org/abs/2005.14165

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of memory and language*, *109*, 104047.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

Campello de Souza, Bruno and Andrade Neto, Agostinho Serrano de and Roazzi, Antonio, Are the New AIs Smart Enough to Steal Your Job? IQ Scores for ChatGPT, Microsoft Bing, Google Bard and Quora Poe (April 7, 2023). Available at SSRN: https://ssrn.com/abstract=4412505 or http://dx.doi.org/10.2139/ssrn.4412505

Caplan, B. (2018). *The case against education: Why the education system is a waste of time and money.* Princeton University Press.

Chalmers, D. J. (2010). The Character of Consciousness. Oxford University Press.

Chalmers, D. J. (2022). Reality+: Virtual Worlds and the Problems of Philosophy. W. W. Norton & Company.

Churchland, P. S. (2013). Touching a Nerve: The Self as Brain. W. W. Norton & Company.

Clarke, A. C. (1998). *3001 The Final Odyssey: A Novel*. Del Rey.

Cotra, A. (2020). Draft report on AI timelines. Open Philanthropy. 2020 Draft Report on Biological Anchors - Google Drive

Cotra, A. (2023). Two-year update on my personal AI timelines. AI Alignment Forum. Two-year update on my personal AI timelines — AI Alignment Forum

Csikszentmihalyi, M. (2008). *Flow: The Psychology of Optimal Experience*. Harper Perennial.

Damasio, A. (2000). The Feeling of What Happens: Body and Emotion in the Making of Consciousness. Mariner Books.

Damasio, A. (2006). Descartes' Error: Emotion, Reason, and the Human Brain. Penguin Books.

Davidson, T. (2023). What a compute-centric framework says about takeoff speeds. Open Philanthropy. What a compute-centric framework says about takeoff speeds - Open Philanthropy

Dennett, D. C. (1991). Consciousness Explained. Little, Brown and Co.

Dennett, D. C. (2005). Sweet Dreams: Philosophical Obstacles to a Science of Consciousness. MIT Press.

Eagleman, D. (2015). The Brain: The Story of You. Vintage.

Ellington, A. J. (2000). *Effects of hand-held calculators on precollege students in mathematics classes: A meta-analysis*. The University of Tennessee.

Ellington, A. J. (2006). The effects of non-CAS graphing calculators on student achievement and attitude levels in mathematics: A meta-analysis. *School Science and Mathematics*, *106*(1), 16-26.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*.

Fiani, B., Reardon, T., Ayres, B., Cline, D., Sitto, S. R., Reardon, T. K., ... & Cline, D. D. (2021). An examination of prospective uses and future directions of neuralink: the brain-machine interface. *Cureus*, *13*(3).

Ford, M. (2015). *Rise of the Robots: Technology and the Threat of a Jobless Future*. Basic Books.

Fuhrmann, M., & Lupu, Y. (2016). Do arms control treaties work? Assessing the effectiveness of the nuclear nonproliferation treaty. *International Studies Quarterly*, *60*(3), 530-539.

Future of Life Institute. (2023). Pause Giant AI Experiments: An Open Letter: Pause Giant AI Experiments: An Open Letter - Future of Life Institute

Gardony, A. L., Brunyé, T. T., Mahoney, C. R., & Taylor, H. A. (2013). How navigational aids impair spatial memory: Evidence for divided attention. *Spatial Cognition & Computation*, *13*(4), 319-350.

Gary Becker (1993) [1964]. *Human capital: a theoretical and empirical analysis, with special reference to education* (3rd ed.). Chicago: The University of Chicago Press.

Giannini, S. (2023). Reflections on generative AI and the future of education. UNESCO. Generative AI and the future of education - UNESCO Digital Library

Giattino, C. & Rosser, M. (2023). AI Timelines: What Do Experts in Artificial Intelligence Expect in the Future. *Our World in Data. https://ourworldindata.org/artificial-intelligence*

Gleick, J. (1999). *Faster: The acceleration of just about everything*. Little, Brown.

Goldman Sachs. (2023). The Potentially Large Effects of Artificial Intelligence on Economic Growth. GS Publishing The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodnani) (gspublishing.com)

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and special education*, *7*(1), 6-10.

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research*, *62*, 729-754.

Graeber, D. (2019). Bullshit Jobs: A Theory. Simon & Schuster.

Gray, P. (2013). *Free to learn: Why unleashing the instinct to play will make our children happier, more self-reliant, and better students for life*. Basic Books.

Gruetzemacher, R., Paradice, D., & Lee, K. B. (2019). Forecasting transformative AI: An expert survey. *arXiv preprint arXiv:1901.08579*.

Hamilton, A., & Hattie, J. (2021). *Not all that glitters is gold: Can education technology finally deliver?* Corwin Press

Hamilton, A., & Hattie, J. (2022). *The lean education manifesto: a synthesis of 900+ systematic reviews for visible learning in developing countries*. Routledge.

Hamilton, A., Hattie, J., & Wiliam, D. (2024). Making Room for Impact: A De-Implementation Guide for Educators. Corwin Press

Hattie, J. (2023). *Visible learning: The sequel: A synthesis of over 2,100 meta-analyses relating to achievement.* Routledge.

Harari, Y. N. (2016). *Homo Deus: A brief history of tomorrow*. Random House.

Harris, A. (2019). Conscious: A Brief Guide to the Fundamental Mystery of the Mind. Harper

Harris, S. (2020). Making Sense: Conversations on Consciousness, Morality, and the Future of Humanity. Harper.

Hembree, R., & Dessart, D. J. (1986). Effects of hand-held calculators in precollege mathematics education: A meta-analysis. *Journal for research in mathematics education*, *17*(2), 83-99.

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. Frontiers in Human Neuroscience, 3, 31

Heyes C. New thinking: the evolution of human cognition. Philos Trans R Soc Lond B Biol Sci. 2012 Aug 5;367(1599):2091-6. doi: 10.1098/rstb.2012.0111. PMID: 22734052; PMCID: PMC3385676.

Hoppe, S., Loetscher, T., Morey, S. A., & Bulling, A. (2018). Eye movements during everyday behavior predict personality traits. *Frontiers in human neuroscience*, 105.

Ishikawa, T., Fujiwara, H., Imai, O., & Okabe, A. (2008). Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of environmental psychology*, *28*(1), 74-82.

Kurzweil, R. (2013). *How to create a mind: The secret of human thought revealed*. Penguin.

Lancy, D. F. (2014). *The anthropology of childhood: Cherubs, chattel, changelings*. Cambridge University Press.

Lieberman, P. (2013). The Unpredictable Species: What Makes Humans Unique. Princeton University Press.

Lluka, T., & Stokes, J. M. (2023). Antibiotic discovery in the artificial intelligence era. *Annals of the New York Academy of Sciences*, *1519*(1), 74-93.

Lowder, S. K., Skoet, J., & Raney, T. (2016). The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World development*, *87*, 16-29.

McAfee, A., & Brynjolfsson, E. (2016). Human work in the robotic future: Policy for the age of automation. *Foreign Affairs*, *95*(4), 139-150.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5, 115-133.

McKinsey. (2023). The economic potential of generative AI: The next productivity frontier. Economic potential of generative AI | McKinsey

Metaculus. (2023). When will the first general AI system be devised, tested, and publicly announced? Date of Artificial General Intelligence | Metaculus

Metzinger, T. (2009). The Ego Tunnel: The Science of the Mind and the Myth of the Self. Basic Books.

Miller, C. (2022). *Chip war: the fight for the world's most critical technology*. Simon and Schuster.

Minsky, M. (2006). The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. Simon & Schuster.

Moravec, H. (1988). Mind Children: The Future of Robot and Human Intelligence. Harvard University Press.

Munro, M. J., & Derwing, T. M. (2001). Modelling perceptions of the comprehensibility and accentedness of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, *23*(4), 451-468.

Nagel, T. (1974). What Is It Like to Be a Bat? The Philosophical Review, 83 (4): 435–450.

Nation, P. (2009). Reading faster. *International Journal of English Studies*, *9*(2).

Nørretranders, T. (1991). The User Illusion: Cutting Consciousness Down to Size. Viking.

Norvig, P. & Russell, S. (2021) – Artificial Intelligence: A Modern Approach. Fourth edition. Pearson.

OpenAI. (2023). GPT-4 Technical Report. https://doi.org/10.48550/arXiv.2303.08774

O'Shea, M. (2005). The Brain: A Very Short Introduction. Oxford University Press. Link

Our World in Data. (2023a). Number of People Working in Agriculture. Employment in Agriculture - Our World in Data

Our World in Data (2023b). Global Literacy Today. Literacy - Our World in Data

Paul, T. V. (2020). *The tradition of non-use of nuclear weapons*. Stanford University Press.

Penrose, R. (1989). The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics. Oxford University Press.

Plato & Jowett, B. (2010) Plato. *Phaedrus*. Vol. 1. Actonian Press

Prinz, J. (2004). Gut Reactions: A Perceptual Theory of Emotion. Oxford University Press.

Prinz, J. (2017). The Conscious Brain: How Attention Engenders Experience. Oxford University Press.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. OpenAI Blog. Language Models are Unsupervised Multitask Learners (d4mucfpksywv.cloudfront.net)

Reagan, T. (1994). Developing "Face and Heart" in the Time of the Fifth Sun: An Examination of Aztec Education. Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994). ED368537.pdf

Rees, M. & Livio, M (2023). Most Aliens May Be Artificial Intelligence, Not Life as We Know It. Scientific American. Most Aliens May Be Artificial Intelligence, Not Life as We Know It - Scientific American

Romer, Paul M. 1994. "The Origins of Endogenous Growth." Journal of Economic Perspectives, 8 (1): 3-22.

Roser, M. (2023). AI timelines: What do experts in artificial intelligence expect for the future? Our World in Data. AI timelines: What do experts in artificial intelligence expect for the future? - Our World in Data

Royal Society. (2019). iHuman Blurring lines between mind and machine. The Royal Society. royalsociety.org/ihuman-perspective

Russell, S., & Norvig, P. (2016). Artificial Intelligence: A Modern Approach. Pearson.

Schleicher, A. (2019). Should schools teach coding? OECD Education and Skills Today, Should schools teach coding? - OECD Education and Skills Today (oecdedutoday.com)

Schumpeter, J. A. (2015). Capitalism, socialism, and democracy. Sublime Books; 2nd edition

Seldon, A., & Abidoye, O. (2018). *The fourth education revolution*. Legend Press Ltd.

Seung, S. (2012). Connectome: How the Brain's Wiring Makes Us Who We Are. Houghton Mifflin Harcourt. Link

Shanahan, M. (2015). The Technological Singularity. MIT Press

Sharma A, Virmani T, Pathak V, Sharma A, Pathak K, Kumar G, Pathak D. Artificial Intelligence-Based Data-Driven Strategy to Accelerate Research, Development, and Clinical Trials of COVID Vaccine. Biomed Res Int. 2022 Jul 6;2022:7205241. doi: 10.1155/2022/7205241. PMID: 35845955; PMCID: PMC9279074.

Solow, R. M. (1956). A contribution to the theory of economic growth. The quarterly journal of economics, 70(1), 65-94.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *science*, *333*(6043), 776-778.

Standing, G. (2017). *Basic income: And how we can make it happen*. Penguin UK.

Tang, J., LeBel, A., Jain, S., & Huth, A. G. (2023). Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience*, 1-9.

Tegmark, M. (2017). Life 3.0: Being Human in the Age of Artificial Intelligence. Knopf.

Terwiesch, C. (2023) Would Chat GPT Get a Wharton MBA? A Prediction Based on Its Performance in the Operations Management Course, Mack Institute for Innovation Management at the Wharton School, University of Pennsylvania Christian-Terwiesch-Chat-GTP-1.24.pdf (upenn.edu)

Turing, A. M. (1950). Mind. *Mind*, *59*(236), 433-460.

U.S. Department of Education, Office of Educational Technology. (2023) Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations, Washington, https://tech.ed.gov/

UNESCO, (2019). World Population Prospects: Highlights. (ST/ESA/SER.A/423). UNESCO.

UNESCO. (2021). AI and education Guidance for policy-makers. UNESCO.

UNESCO. (2022). Recommendation on the Ethics of Artificial Intelligence. UNESCO.

UNESCO. (2023). ChatGPT and Artificial Intelligence in Higher Education Quick start guide. UNESCO.

United States. (1966). *National Commission on Technology, Automation, and Economic Progress. Automation and Economic Progress.* U.S. Government Printing Office.

Van Parijs, P. (1995). *Real freedom for all: What (if anything) can justify capitalism?*. Clarendon Press.

Van Parijs, P., & Vanderborght, Y. (2017). *Basic income: A radical proposal for a free society and a sane economy*. Harvard University Press.

*Warneke, B.; Last, M.; Liebowitz, B.; Pister, K. S. J. (January 2001). "Smart Dust: communicating with a cubic-millimeter computer". Computer.* ***34*** *(1): 44–51. doi:10.1109/2.895117. ISSN 0018-9162.*

Yang, A. (2018). *The war on normal people: The truth about America's disappearing jobs and why universal basic income is our future*. Hachette UK.

Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. *Global catastrophic risks*, *1*(303), 184.

Zhang, B., Dreksler, N., Anderljung, M., Kahn, L., Giattino, C., Dafoe, A., & Horowitz, M. C. (2022). Forecasting AI progress: Evidence from a survey of machine learning researchers. *arXiv preprint arXiv:2206.04132*.