EDUCATION

# WELCOME TO THE MACHINE

## Opportunities and risks of generative Artificial Intelligence for education

**Michael Johnston**
**Foreword by Oliver Hartwich**

THE NEW ZEALAND INITIATIVE

www.nzinitiative.org.nz

THE NEW ZEALAND INITIATIVE

**Published May 2024 by**
The New Zealand Initiative
PO Box 10147
Wellington 6143
New Zealand
**www.nzinitiative.org.nz**

Views expressed are those of the author and do not necessarily reflect the views of The New Zealand Initiative, its staff, advisors, members, directors or officers.

**ISBN**
978-1-7386277-0-7 (print)
978-1-7386277-1-4 (online)

RR80

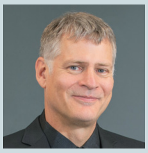
Printing arranged by TBD Digital

# WELCOME TO THE MACHINE

## Opportunities and risks of generative Artificial Intelligence for education

**Michael Johnston**
**Foreword by Oliver Hartwich**

## ABOUT THE AUTHOR



Dr Michael Johnston is a Senior Fellow at The New Zealand Initiative. He leads the workstream on education. Before joining the Initiative, Dr Johnston held academic positions at Victoria University of Wellington (2011–22), including Associate Dean (Academic) in the University's Faculty of Education (2020–22). He was a lecturer in psychology at the University Melbourne and a Research Fellow at Latrobe University. Dr Johnston holds a PhD in Cognitive Psychology from the University of Melbourne.

## ACKNOWLEDGEMENTS

I am grateful to Maciej Surowiec for giving me the initial idea for this report, and for fascinating discussions along the way. Many thanks also to Vanessa Sorenson and Microsoft New Zealand for hosting a forum on AI in Education in June 2023 and to all participants in the forum. That discussion sparked many of the ideas reflected in this report. I am also grateful to Serge van Dam for testing my ideas in a stimulating conversation.

# Contents

# Foreword

I must confess: I am a massive nerd and an even bigger techie. Ever since I got my hands on a Commodore Amiga 500 back in 1987 (with a whopping 512kb of RAM!), I have been hooked on technology.

Over the years, I have tried nearly every new gadget and gizmo that has come along. Call me a techno-optimist, but I cannot help but be amazed by the incredible advances I have seen in my lifetime. And I am only 48! It is mind-boggling to think about how far we have come since the days of my trusty Atari console in the early 1980s.

So when it comes to the rapid rise of Artificial Intelligence (AI), I am filled with a sense of excitement and possibility. As someone who loves to tinker and explore, I am constantly turning to my favourite AI tools to ask questions and learn new things. It is like having a personal tutor available around the clock to satisfy my curiosity about quantum physics or help me finally grasp complex economic concepts. The results are often simply stunning.

But here is the thing: My ability to effectively harness AI stems from the solid educational foundation I have built over the years. It is this base of knowledge that allows me to ask the right questions, interpret the answers, and yes, even spot the occasional bit of AI-generated nonsense.

And that is why, even as an AI enthusiast, I am growing increasingly concerned about its implications. Like other transformative technologies before it, AI has the potential to be a great polariser. Those who are already knowledgeable and educated will be able to wield it as a powerful tool to augment their intelligence and abilities. But for those without a strong foundation to build upon, AI could simply reinforce their gaps in understanding, leaving them at a disadvantage.

This is where Dr Michael Johnston's insightful new report, 'Welcome to the Machine,' comes in. As a cognitive psychologist, Michael deeply understands the inner workings of the brain and the science of learning. While he may not share my unbridled technophilia (much to his bemusement), he recognises AI's potential to enhance education when used judiciously and with clear purpose.

Michael's report is a timely guide for navigating this complex new landscape. He presents a compelling case for why a solid educational foundation is more critical than ever in an AI-powered world. Students must still master core skills and knowledge, even as AI offers tantalising shortcuts. At the same time, he highlights the ways in which AI can be a powerful ally for teachers and learners when thoughtfully applied.

In essence, Michael points us toward a balanced approach that combines technological enthusiasm with healthy scepticism. It is a conclusion he and I have often reached in our spirited discussions, even if we have taken different paths to get there.

I wholeheartedly commend this lucid and thought-provoking report to anyone with a stake in the future of education: policymakers, educators, parents, students, and yes, even the AI chatbots themselves. Michael has done us all a great service by illuminating the path forward.

And now if you will excuse me, I have some questions about string theory to pose to my AI physicist friend...

**Oliver Hartwich**
Executive Director, The New Zealand Initiative

# Executive Summary

In this report, the role of artificial intelligence (AI) in education is analysed using a science-based account of human learning. The report challenges the notion that AI will alter what is important for students to learn or the way they learn from a cognitive perspective.

An important task of an education system is to inculcate disciplinary knowledge. AI will not change that. Disciplinary knowledge such as science, history and mathematics provides crucial epistemological tools; that is, methods for testing truth claims. Human progress in knowledge would be impoverished without scholars conversant in these disciplines. Moreover, acquiring at least rudimentary knowledge of epistemic disciplines supports young people in becoming competent democratic citizens. Democracy relies on citizens capable of critical thinking. Critical thinking itself relies both on knowledge and an ability to reason. Both are nurtured by disciplinary learning.

Research in cognitive psychology suggests that teachers' practice can benefit from a practical understanding of working memory and cognitive load. Working memory is a short-term human memory system that mediates the performance of new tasks and procedures. As such, it plays a central role in learning numeracy, mathematics, literacy, science, and other academic disciplines. Working memory has a very small storage capacity, and its contents are quickly forgotten unless rehearsed. Cognitive load refers to the working memory resources required to perform a task; a high-cognitive-load task occupies most or all the capacity of working memory.

The limitations of working memory are such that students must not rely on technology to perform cognitive operations before they are reliably encoded in long-term memory. If they do, those operations will remain mediated by working memory, and any further learning that relies on them is likely to cause cognitive overload.

To think critically, students must have knowledge. If we do not have a store of knowledge and concepts at our immediate disposal, thinking is empty. It is not enough to be able to quickly find information online. The cognitive structures that are established by knowing and understanding are indispensable to criticality and creativity. Similarly, writing is one of the greatest technologies for the enhancement of thinking ever invented. By enabling us to 'outsource' our thoughts to text, writing vastly increases the amount of immediate-term complexity we can cope with.

AI will not, therefore, obviate the need for students to learn to acquire knowledge and writing skills for themselves, irrespective of its capacity to collate knowledge and to produce writing much better and more efficiently than most students (and most adults).

A more general principle for the use of any technology in education emerges from these observations: Before relying on technology for knowledge and skills prerequisite to later learning, it must be securely encoded in long-term memory.

## CHAPTER 1
# Introduction

The emergence of generative large language models has sparked considerable discussion on its implications for education. Initially, concerns focused on assessment: Some educators have expressed anxiety that generative artificial intelligence (AI) compromises confidence in the authenticity of students' work.[1] For example, AI engines can write nearly flawless prose. Any assessment intended directly to measure students' writing skill would therefore be rendered meaningless by their unrestricted use of AI chatbots.

Discussion quickly shifted from the potential for AI to enable cheating in summative assessment to its potential impact on teaching and learning. Here, there are potential opportunities as well as risks. Depending on what it is used for, AI may improve or hinder student's learning. Similarly, it may enhance or degrade teachers' practice.

AI tools are likely to become more and more powerful. How they will develop and the functions of which they will become capable are unpredictable. Faced with this unpredictability, some educational authorities have adopted an extreme position of attempting to ban or avoid the use of AI engines in education.[2] That is not realistic – AI is already being used widely and bans are nearly impossible to enforce. Neither is it an optimum position, even were banning AI possible. Used appropriately, AI stands to enhance teaching and learning. Another extreme position would be to adopt a laissez-faire approach, allowing students and teachers to use AI in an unfettered manner. That would not be beneficial for education either.

A principled and evidence-based framework is required to guide the educational use of AI as it continues to increase in functionality. Such a framework would help maximise the positive contributions of generative AI for teaching, learning and assessment, and mitigate its risks.

In this report, the potential for generative large-language models to contribute to education, as well as its associated risks, are discussed. Particular attention is paid to human cognitive architecture and its implications for teaching and learning. The constraints these considerations place on the appropriate use of large language models (and indeed any technology) in education are explored. As we will see, from a cognitive perspective, it will remain important for young people to learn to read and write, and to become numerate. Despite the existence of AI technology and online resources, it will also remain important for them to have knowledge and to grasp concepts.

The report lays out a framework for identifying ways in which AI can enhance cognitive and epistemic development, and to avoid uses that may undermine it. No attempt is made to provide an exhaustive account of the possibilities for involving AI in education. Indeed, because generative AI is nascent and will develop in unpredictable ways, no such account would be possible. Rather, the report suggests principles for assessing whether an educational application of AI is likely to have positive or negative consequences for students' learning. While the focus is specifically on AI, the analysis is, in many ways, generalisable to the use of other technology in educational contexts. The report is intended to offer insights to policymakers, teachers and parents, to guide the use of AI in ways that are educationally beneficial.

Generative AI can be of direct benefit in three core, overlapping educational areas: Support for learning, support for teaching, and formative assessment.

Support for learning includes the possibility of AI as a tool to enhance students' own cognitive activities. Students might use AI to help them gather information, edit documents, or check the technical details of their writing. Support for teaching includes the use of AI as a 'virtual tutor' and to gather and analyse data to improve teachers' practice. Formative assessment – from brief, informal interactions between students and teachers to structured feedback on substantial learning activities – is a powerful pedagogical tool.[3] While the involvement of AI in assessment poses a challenge to assuring authenticity in summative assessment, it has considerable potential to support and enhance formative assessment.

CHAPTER 2

# Generative AI and human cognition

Cognitive psychology is the science of human perception, memory, attention and language. It is relevant to this report in two ways. First, it provides a scientific account of human learning, with clear applications to formal education.[4] As such, it has much to contribute to discussions about ways in which AI can be used to enhance education and ways in which it may inadvertently undermine it. Second, the computational architecture underpinning generative AI has its origins in models of human cognition.[5]

Cognitive psychology has generated a large body of theory explaining human cognitive capacities. While many fundamental issues remain unresolved, these theories have advanced enough to have powerful applications. An important application relevant to this report is the 'science of learning'. The science of learning uses cognitive psychology to inform effective teaching practice. Importantly, the cognitive theories contributing to the science of learning are all computational – they recognise that learning involves automating cognitive operations.

## 2.1 Computational architecture

The similarities and dissimilarities of AI and human brains in the ways each processes information is an essential consideration in discussions about the appropriate uses of AI in education.

A long-running debate in cognitive psychology concerns the extent to which human information processing is *modular*. Modular cognitive processes are specialised for specific tasks. For example, the recognition of written words by skilled readers is a modular process. Cognitive modules have features that make them highly efficient, an essential property for skilled performance.[6]

Modules are *automatic*, meaning that they always operate in the presence of an input matching their specialised domains of operation. They are also *encapsulated*, meaning that their operations are insulated from those of other cognitive systems. This means that they cannot be interrupted or slowed down by distracting stimuli or thoughts. Theories that explain human cognition in terms of modular processes are known as *computational* theories.

An example of a modular process is the visual recognition of written words by skilled readers. A skilled reader cannot attend to a written word without reading it. This is demonstrated by the Stroop effect.[7] In Stroop's experiment, skilled readers were shown written colour names and asked to report their font colours. They were slower to do this when the colour names and font colours did not match (e.g., blue), than when they did (e.g., red). The Stroop effect shows that a skilled reader is unable to avoid reading a word even when it harms the performance of a different task involving that word (e.g., naming its font colour).

Beginning readers are no slower to report font colours that are incongruent with written colour names than they are with congruent ones. That is because they have not yet formed a cognitive module for recognising written words. In other words, the process of reading colour names is not yet automatic for them, as it is for skilled readers. We will return to the importance of cognitive automaticity to human learning in the next section.

Some cognitive theorists – the 'connectionists'[8] – rejected modularity and computational theories of cognition. Instead, they pursued a line of modelling known as parallel distributed processing (PDP). PDP models are simulated networks of interconnected nodes arranged in layers, bearing a superficial resemblance to the arrangement of neurons in the brain. For this reason, they have sometimes been called 'neural networks.' Early models typically had just three layers: An 'input' layer, encoding a stimulus as a pattern of activation across its nodes; a 'hidden' layer; and a layer encoding output.

In recent decades, exponential increases in the processing speed and storage capacity of digital computers have led to a massive expansion in the scale of connectionist models. Early PDP modellers developed networks that simulated such cognitive functions as word recognition[9] and human-like temporal effects in the processing of sequentially presented stimuli.[10] Over time, they became more and more impressive in what they can simulate. Eventually, this line of modelling gave rise to generative AI, including large language models. Large language models have similar architecture to the early connectionist models, but with many more layers. For example, the largest version of GPT-3 has 175 billion nodes arranged in 96 layers.[11]

Connectionism, then, has given us generative AI, while computational theories of cognition have given us tools to develop a critical approach to its use in education. The science of learning offers a framework within which to understand the potential ways in which AI can impact human learning, for better and for worse.

### How large language models work

As we have noted, large language models, like all PDP models, are implemented on virtual networks comprising nodes arranged in layers. Input text – comprising a prompt – is first divided into groups of characters, which may be words, prefixes, suffixes or punctuation marks. Each such group is called a *token*. Each token is assigned a unique number. The sequence of token values corresponding to a prompt is the input to the model. It is encoded in the topmost (input) layer of the network by assigning the token values to nodes.

Each node in the input layer is connected to every node in the first hidden layer. Each node of the first hidden layer is similarly connected to every node in the second hidden layer, those of the second layer to those of the third, and so on, through to the final (output) layer. The values of the output layer are, once again, numbers corresponding to tokens. Across the nodes of the output layer, these values are recoded as text – the response of the network to the prompt.

The strengths of the connections between nodes are governed by values called *weights*. The configuration of the weights governs how the network responds to a given prompt. Weight values are initially set by 'training' the network on a large corpus of text. This comprises the learning capability of the network. Regularities in the training corpus are encoded in the network by the training process. Following training, it uses this information to respond to prompts by predicting the word (or other character string) most likely to be next in an output sequence. Networks can continue to learn as they interact with users by adjusting the weights in the system in response to ongoing feedback.

## 2.2 Differences between human language and large language models

It can be tempting to believe that generative AI models process language in a human-like way, simply because they produce uncannily human-like and frequently very useful responses to plain language prompts. Furthermore, the arrangement of AI virtual networks, in multiple layers of nodes, resembles the arrangement of neurons in a human brain, superficially at least. The impression that large language models process information in a human-like way is, however, illusory. The ways in which humans and AI arrange the symbols of language (grammar) and their relationships with the meaning of language (semantics) are fundamentally different.

The capacity to represent objects, events and personal experience symbolically makes human language possible. Linguists call the units of meaning in language *morphemes*. Morphemes are most often words, but also include prefixes (un-, dis-, pre-, etc.) and suffixes (-ing, -ed, -ly, etc.). The rules by which morphemes may be combined to express different meanings comprise the grammar of a language.

The founder of the discipline of psycholinguistics, Noam Chomsky, argued that human cognitive architecture is equipped with an innate and universal grammar, which simultaneously constrains and renders powerful the grammatical structure of human language.[12] Other theorists have disputed this claim.[13] Whether Chomsky is correct or not, the grammatical, rule-governed way that human language combines symbols to construct meaning is very different than the way large language models produce their output.

Although they can produce grammatically flawless text, large language models have no way of representing the rules of grammar. Writing with fellow linguist Ian Roberts and AI specialist Jeffrey Watumull, Chomsky has pointed out that, as powerful as they are, AI chatbots cannot cope with certain grammatical structures that are trivial to human beings.[14] Chatbots have become very adept at mimicking human language, but they do not process language information in anything like the way the human brain does.

Even more profound than the difference in the way humans and large language models produce grammatical sentences, are their differences regarding meaning. At no point do large language models refer to the meanings of words. Indeed, they do not encode meaning at all. Instead, they predict output based on which tokens and combinations of tokens are likely to be proximal in text.

Some theorists, from Dreyfus in 1965 to Fjelland in 2020, have argued that connectionist models cannot, even in principle, learn to use language in a human-like way, because they are not *embodied* as human beings are.[15] In being embodied, human beings are no different than any other biological organism. But, unlike other animals, human beings have also evolved cognitive systems for mentally representing and communicating about their environments, and their own experience, *symbolically*; that is, by combining morphemes to express meaning using grammar. The grammar of language enables us to construct an infinity of novel meanings. In this way, humans are unique.[16]

Because human beings are embodied, human language is grounded in a world of objects and events which we perceive, and with which we interact. The physical embodiment of human beings – our ability to interact with objects in the environment to meet our biological and psychological needs – provides a basis for the mental representation of meaning. Language enables us to express and communicate that meaning symbolically.

Unlike human beings, large language models do not need anything from the world, so do not act with the purpose that derives from biological drives.

They do not, therefore, have any basis for using language as a means of communicating meaningfully. What they produce may be meaningful to human beings, but – lacking both bodies and inner life – they do not experience meaning. Rather, they *simulate* the meaningful use of language based on the statistical co-occurrence of words and phrases. They do not *know* anything and will assert falsehoods when the strings of words comprising those falsehoods correspond to the statistical likelihoods distributed throughout a network's nodes. Large language models cannot 'fact check' their own statements because they have no conception or experience of truth or falsity.

The embodiment problem cannot be solved simply by putting large language networks inside robots with the ability to sense and manipulate objects. The concept of embodiment goes deeper than that. Human bodies have evolved over countless generations through complex interaction with a constantly changing environment. They perceive, navigate and manipulate the environment to fulfil biological drives and motivations. It is the *affordances* of objects and processes to satisfy our drives and motivations that provide the ground of meaning for human language.[17] Gibson provides an informative example of what *affordance* means in the context of embodied organisms:

> If a terrestrial surface is nearly horizontal (instead of slanted), nearly flat (instead of convex or concave), and sufficiently extended (relative to the size of the animal) and if its substance is rigid (relative to the weight of the animal), then the surface affords support. … It is stand-on-able, permitting an upright posture for quadrupeds and bipeds. It is therefore walk-on-able and run-over-able. It is not sink-into-able like a surface of water or a swamp, that is, not for heavy terrestrial animals. Support for water bugs is different.
>
> — J.J. Gibson[18]

A corollary of the incapacity of large language models to represent meaning is that they cannot develop a 'theory of mind.' This term refers to the ability of human beings to ascribe intentional states to others.[19] In plain terms, this means that we can conceptualise the emotional, motivational and cognitive states of others well enough to make human society possible. Teachers are unlikely to be effective if they do not understand the intentional states of students; effective teachers must have accurate theories of mind. Because generative AI models lack this capacity, they cannot, therefore, substitute for skilled teachers.

These fundamental differences between human beings and large language models have important implications for the applications of AI in education. However, this does not imply that large language models lack educational value. Rather, they should be treated as tools under the control of skilled human teachers. Used in this way, large language models offer great potential to enhance teachers' productivity.

CHAPTER 3

# AI as a support for learning

In the mid-1970s, the advent of affordable hand-held calculators provoked a debate in educational circles that, in some ways, mirrors current debates about generative AI. Many educators saw benefits for students' motivation to learn mathematics if they were able to concentrate on the application of mathematical concepts to 'real world' problems, without getting bogged down in the mechanics of mental arithmetic.[20] Others feared that, if students came to rely on calculators to do their arithmetic for them, they would not learn basic computational skills. Later research has shown this fear to be justified.[21]

Around the same time as the calculator debate, research in cognitive psychology was making rapid advances. It would be another decade, however, before that body of research was mature enough to be applied to teaching practice, through what is now known as the science of learning. Had the science of learning been as advanced then as it is now, it would have informed the calculator debate. It can certainly inform the current debate regarding the role of AI in education.

## 3.1 Working memory

The theory of *working memory* has made an especially important contribution to the science of learning. The term was coined by Pribram, Miller and Galanter in 1960, but much of the early theoretical and empirical development was carried out by English psychologist Alan Baddeley and his colleagues.[22] Baddeley's research team built on Atkinson and Shiffrin's 1968 model of human memory.[23] They theorised working memory as a short-term memory store underpinning the conscious, immediate-term processing of information. The model comprises three subsystems: A phonological loop, which stores speech information; a visuospatial store, which stores information about the shapes and locations of objects; and an episodic buffer, which holds information recalled from long-term memory. The information stored in these subsystems is available for the conscious performance of cognitive operations.

Decades of research followed, testing and refining the theory. The theory of working memory is now supported by a large volume of empirical data. It forms an important component of understanding the cognition of human learning. That understanding, in turn, has important implications for teaching.

Working memory stores information while we use it to perform cognitive tasks. For example, when we perform mental arithmetic, working memory is used to store the numbers we are operating on, as well as any interim solutions. Working memory takes input both from the sensory environment and from long-term memory. In this way, it serves to integrate novel and stored information.

Working memory has two salient limitations: It has a very small capacity, storing as little as four items of information;[24] and it decays within a few seconds unless it is actively maintained through rehearsal. For example, if someone dictates their telephone number, you must repeat it to yourself until you can record it to avoid forgetting it. Such repetition maintains information in the phonological loop of working memory. So, while working memory supports powerful cognitive functioning, it does not retain information for long, and is easily overloaded.

For durable learning to occur, information must be transferred from working memory to long-term memory, which has an effectively infinite capacity. Indeed, in cognitive terms, learning has been *defined* as the transfer of knowledge and processes from working memory to long-term memory.[25]

The term *cognitive load* refers to the demands a task imposes on working memory. When task demands exceed the capacity of working memory, performance suffers. In educational settings, this is especially likely in cognitively demanding learning such as mathematics, science and early literacy. Cognitive load is therefore a particularly important consideration in instructional design.[26] Teachers must be able to manage the cognitive load students experience while they are assimilating new concepts and skills. Failure to do so results in cognitive overload and failure to learn. Cognitive overload may be accompanied by feelings of confusion and, if the state persists for too long, loss of learning efficacy and demotivation.[27]

## 3.2 Strategic and automatic cognition

Cognitive psychologists distinguish between *strategic* and *automatic* cognition. Strategic cognition is mediated by working memory. It is slow and deliberate yet highly flexible and able to draw in information from many sources. Strategic cognition is subject to the limitations of attentional bandwidth, and of the small capacity and rapid decay of working memory. Automatic cognition is mediated by long-term memory and is modular in nature (see 2.1). This means that automatic processes only function when input information aligns with a template; they operate without our awareness; and do not need focused attention.

The transfer of information from working memory to long-term memory may be conceptualised as a process of *automation*. A commonplace example of automation is learning to drive. At first, this is a slow and difficult task, involving the coordination of many sub-tasks, such as using the brake and accelerator, steering, indicating, and monitoring the road for curves and other vehicles. With practice these sub-tasks become automated and less and less concentration is required. Experienced drivers can converse with passengers without loss of performance, provided driving conditions are not unduly complex. Similar processes of automation apply to learning to play musical instruments, the acquisition of skills in sport and many other performance domains. They apply equally to cognitive skills, including reading, numeracy and disciplinary learning. Indeed, all of what evolutionary psychologist David Geary has termed *biologically secondary* knowledge is acquired in this way.[28] Biologically secondary knowledge is knowledge that human beings have developed culturally. It contrasts with biologically primary knowledge, which is knowledge we have evolved to acquire naturally, often through social interaction. Learning biologically secondary knowledge relies on working memory whereas learning biologically primary knowledge does not.

The distinction between biologically primary and biologically secondary knowledge is illustrated by differences in the way oral language and literacy are acquired. Oral language is biologically primary: All human cultures have oral language; it is acquired without apparent effort during early childhood; and it does not have to be explicitly taught. Literacy, on the other hand, is biologically secondary: Historically, few cultures have been literate; acquiring literacy is a difficult cognitive task; and it requires explicit teaching.

## 3.3 The acquisition of biologically secondary knowledge

Early reading is very slow and deliberate because it is mediated by working memory. With practice, its various subprocesses are encoded in long-term memory and are thereby automated. A child first starting to 'sound out' a word from its spelling must segment it into graphemes (individual letters or letter sequences corresponding to phonemes – e.g., 't', 'ch', 'air'), then assign the correct phoneme to each grapheme, and finally, blend the phoneme sequence. On the other hand, for a skilled reader, visual word recognition is so rapid as to be effectively instantaneous. In other words, for skilled readers, the process of reading has become automatic and working memory is no longer required; however, working memory is still involved in temporarily storing the meaning of what is being read.

In education, it is essential that any knowledge upon which later learning will depend is automated before approaching the more advanced material. If it is not, it will continue to occupy limited working memory resources leaving them unavailable to accommodate the new learning. In this respect, working memory may be conceptualised as a learning 'bottleneck.'

If we allow students to rely on technology for knowledge or skill before they have learned it to the point of automaticity, they will be unable to advance their learning effectively. Technology should never, therefore, be a substitute for any learning upon which later learning depends. This is the case even for skills that might seem redundant in the information age if those skills are beneficial for cognitive development.

The earlier example of calculator technology clarifies this point. To learn calculus or statistical theory – both of which have many applications, especially in science and engineering – students must understand basic algebra. Algebra, in turn, relies on knowledge of ratios (fractions) and balanced equations, among other things. Developing sound understanding of ratios and balanced equations is built on understanding of elementary arithmetic. If young children are allowed to use calculators for arithmetic before achieving cognitive automaticity with those skills, they will often struggle to access higher-order mathematical skills and knowledge. That is because, without automating arithmetic, learning that relies on it is likely to cause cognitive overload. Even basic adult numeracy skills, such as managing personal finance, require conceptual knowledge of percentages (to understand the implications of changes in home-loan interest rates, for example). Again, without having automated basic arithmetic, most students will struggle to grasp what a percentage is.

The use of large language models in education has similar implications for the teaching of writing. Writing is a powerful tool for the development of critical thinking and creativity. However, critical thinking and creativity, by definition, both require innovation, which, again by definition, cannot be automated and will always draw heavily on working memory resources. Thus, original thinking in either critical or creative form is always a cognitively demanding activity. For that reason, the technical skills of writing – from holding a pencil, to correct spelling and sentence construction, to basic compositional skills – must be automated if writers are to have the working memory capacity required to use writing to support critical and creative thinking. If we allow students to use AI to produce texts before they have automated the necessary skills and knowledge to produce them for themselves, we risk depriving them of that powerful support.

## CHAPTER 4

# Current thinking on the uses of generative AI in education

Consideration of the implications of generative AI for education is in its infancy. At the time of writing, it is little more than a year since the initial launch of the most famous large language model, ChatGPT. Two contrasting philosophies of education have emerged in discussions about, and development of, AI in education. One is a social constructivist, '21st-century-learning' paradigm. The other is a cognitive paradigm, founded on scientific understanding of human learning. These positions are not usually made explicit by commentators. Rather, the contrasting philosophies are typically implicit in their assumptions about what young people need to learn, and how they learn.

In this section, these epistemic and pedagogical assumptions are discussed. The social constructivist 21st-century-learning paradigm is illustrated by much of the discussion in a webinar hosted by EdTech New Zealand, on 22 February 2023.[29] The cognitive paradigm has informed the development of *Khanmigo*, a powerful, purpose-specific AI tool for education from the Khan Academy.[30]

### 4.1 EdTech webinar

One of the hallmarks of 21st-century-learning is a strong emphasis on creativity and critical thinking as generic skills. This perspective featured strongly in comments made during the EdTech webinar. It is also evident in much of the broader commentary on AI in education.

According to Kirsty Chadwick, CEO of The Training Room Online (a learning advisory, content development and technology service)[31] a potential function of AI is to teach students to "ask great questions", prompting large language models to produce output that will elicit their critical thinking.[32] Dave Moskowitz,[33] a technology entrepreneur, commented that "it's never been more important to teach kids critical thinking"[34] and that we should "let the machines worry about the facts and the storage."[35] Moskowitz believes that students should learn to use AI to draw together information from many sources and "compose that into new constellations of ideas."[36]

Developing capabilities to think critically and to be creative are indeed important purposes of any education system, but these capabilities are not generic. Critical thinking in science is different than critical thinking in history or the arts. A mechanic diagnosing a problem with a car's engine is thinking critically, but the success of that thinking depends on sophisticated knowledge of mechanics. From a cognitive perspective, knowledge and conceptual understanding are essential components of sophisticated critical thinking.

Working memory is centrally involved in any critical or creative thought. Knowledge and skill are the raw materials from which critical and creative thinking emerge. That knowledge must be robustly encoded in long-term memory and available for automatic recall, leaving free the attentional and mnemonic resources of working memory to support critical and creative thinking.

From a cognitive perspective, original thinking, whether creative or critical, inevitably involves a degree of randomness. Sweller and colleagues called this the "randomness as genesis principle".[37]

According to these theorists, the random generation of ideas makes innovation possible. However, randomness on its own is unlikely to result in anything useful. Randomly generated ideas must undergo a process of selection through existing cognitive frameworks of already-held knowledge, and then integrated into those frameworks. Integration involves reorganising existing frameworks to accommodate a new idea or concept. So, before young people can productively think critically or creatively, they must first have assimilated frameworks of knowledge. Disciplines like science, mathematics and history provide such frameworks, comprising facts, concepts and processes to test new ideas. All these things need to be encoded in long-term memory to support criticality and creativity.

The conceptualisation of critical thinking and creativity as abstract, generic skills is just one way in which a technological-utopian perspective on the potential role of AI in education overlooks insights from the science of learning. Several participants in the EdTech webinar made suggestions for using AI in ways that would risk undermining the learning of foundational knowledge and skills. Francis Valentine, CEO and founder of The Mind Lab,[38] said that we should move away from education based on "rote learning and rote answering."[39] But repetition is essential to encoding knowledge and skill in long-term memory – and repetition needn't be boring if it is embedded in interesting activities.

Madelaine Newman, Executive Director of AI Forum NZ,[40] commented that AI will help us create content in the same way that calculators help us with calculations. Professor Ian Watson, a computer scientist at the University of Auckland also compared the effect of AI on teaching with that of calculator technology. He commented that we can rely on software to correct our writing. This might be taken to imply that learning to write independently will be made redundant by AI. In fact, though, the introduction of calculators – and the subsequent deemphasis of ensuring that students automate arithmetic skills – cautions against adopting AI at the expense of learning fundamental skills like writing.

Data from the Programme for International Student Assessment (PISA) show that, since the first round of PISA testing for mathematics in 2003, the mathematical skills and knowledge of 15-year-old New Zealanders have declined by the equivalent of about one-and-a-half years of schooling.[41] While the proximal cause of the decline in mathematics skills in young people is probably ineffective teaching,[42] overreliance on calculators may also have had an effect. Indeed, the kind of pedagogy that has accompanied the decline in young people's numeracy – including an emphasis on teaching multiple strategies to solve problems, rather than on automating basic arithmetic – was arguably itself enabled by calculator technology.

While Newman and Watson are almost certainly correct in their respective views on the potential impact of AI on content creation and teaching practice, the caveats that apply to the use of calculators for arithmetic also apply to using AI. Allowing young children to use calculators before they have achieved sufficient cognitive automaticity with arithmetic skills inhibits the development of those skills.[43] This, in turn, decreases their chances of being able to acquire more advanced mathematical skills.[44] Similarly, allowing students to use AI to create text before they have automated technical writing skills would inhibit their capability to use writing as a tool of thinking.

Writing enables students to partially overcome the limitations of working memory. Skilled writers can commit their ideas to text, which acts as a record of those ideas, obviating the need to store them in working memory while further developing their thinking. Moreover, the process of editing can be seen as a process of refining ideas. Failing to teach young people technical writing skills would therefore deprive them of powerful tools to develop their thinking.

Allowing large language models to produce ideas for students is no substitute for enabling them to develop their own thinking skills through writing.

Francis Valentine emphasised that AI would bring to the fore education in 'soft skills' because these "won't be touched" by AI.[45] Again, this is consonant with the so-called 21st-century-learning paradigm, which often evokes the impact of technology on the employment environment to justify radical shifts in education. While education must indeed keep pace with technological change, the fundamental structure of human cognition – including the neural mechanisms of learning and attention – do not change with technological development.

Not all participants in the EdTech webinar took a utopian view of AI in education. Dave Moskowitz referred to ChatGPT as a "Dunning-Kruger machine".[46] This references the 'Dunning-Kruger effect', a cognitive bias whereby people of little competence in a particular domain tend to over-estimate their competence. Moskowitz clarified his comment by noting that large language models produce their output purely stochastically – as discussed earlier in this report – but typically do so with a tone of great confidence. He further noted that sessions that go on for a long time can result in the models going down "weird rabbit holes."[47]

Moskowitz's warning is a useful counterpoint to a claim from Francis Valentine that AI will "be able to identify what is real and what is not", and therefore be useful for curbing disinformation.[48] As Moskowitz pointed out, the 'knowledge' reflected in the responses of large language models is not, in fact, knowledge at all. The quality of the information produced in the responses of large language models is only as good as that of the text used in the training corpus. Moskowitz gave a documented example of a language model asserting that a kilogram of steel is heavier than a kilogram of feathers. One of the model's statements in this response was, "The exact weight of a kilogram of feathers would depend on the size and type of the feathers, but it would generally be significantly less than a kilogram."[49] Responses like these demonstrate that the fact-checking capacity of large language models is dubious.

Members of the EdTech panel gave some useful perspectives on the impact of AI on educational assessment. Professor Watson noted that AI chatbots currently score about 65% on his university exams in computer science. This is impressive, especially considering the early stage of AI technology. In light of the potential for cheating using large language models in uninvigilated assessments, Watson believes that supervised exams will make a comeback.

Other panellists argued that, rather than returning to the ubiquitous use of rigid examinations, AI could open up new approaches to assessment and credentialling. Kevin Bell, Head of Research and Education at AWS (Amazon Web Services) commented that AI provides an opportunity to "revisit learning outcomes" – by which he seems to have meant an opportunity to develop new assessment formats.[50] He argued that, if an interaction with an AI tool can certify that a student has met a learning outcome (or assessment criterion), there may be no need for an examination. Moskowitz believes that teaching, practice, assessment and credentialling can be treated as essentially the same thing. In his view, we have moved on from the necessity to write essays to demonstrate learning. This is an interesting perspective, which will be explored later in this report.

Moskowitz also noted opportunities for the personalisation of teaching. AI engines can use a student's responses to select material appropriate to his or her level of understanding. A related function is to give students practice on examples tuned to their levels of ability. AI can do this in the same manner as computer adaptive tests, which have existed for some time. Moskowitz also described existing AI platforms for coaching students in writing, translating between languages, tutoring in mathematics, and a virtual medical coach that enables student doctors to practice surgery on virtual patients.

## 4.2 Khanmigo

In 1984, educational psychologist Benjamin Bloom wrote on what he called the *two-sigma problem*.[51] Bloom reported data showing that one-on-one tutoring could improve educational achievement by a full two standard deviations (sigma is the Greek letter typically used to denote one standard deviation; hence 'two-sigma'). In plain terms, a two standard deviation improvement means that an average student given one-on-one tutoring performs at the 98th percentile on the distribution of achievement for students not receiving tutoring. The 'problem' is that providing every child with a one-on-one tutor would be prohibitively expensive.

Sal Khan, of the Khan Academy, believes that AI may offer a solution to the two-sigma problem. While acknowledging the risks AI poses in education, Khan pointed out that ignoring the implications of AI for education also carries risks. Apart from forgoing potential benefits to learning, ignoring or attempting to ban AI would cede the ground to 'bad actors' marketing products that may harm education. His solution was to develop a specific-purpose education AI, called Kahnmigo, to maximise the benefits and minimise the risks of the technology.[52]

Khanmigo operates as a personal tutor and teaching assistant. It is designed to be used with the oversight of a human teacher. Khan described some of its functions and features in a recent TED talk.[53] Kahnmigo was released early in 2023 but, at the time of writing, it is available only in the United States.

Khanmigo tutors students in many domains of learning. For example, it can present students with algebra problems and, when they make errors, ask them to explain their reasoning processes to understand where they have gone wrong. It can remind students of rules they appear to have forgotten to apply and ask them to try a problem again with those rules in mind. This approach builds understanding in a way that simply supplying students with correct answers would not.

By ensuring that students first try to solve problems and develop their own thinking, rather than doing these things for them, Khanmigo avoids many of the risks that generative AI might otherwise present to students' learning.

A striking feature of Khanmigo is that it 'coaches' itself about how to tutor, using a parallel AI. When Khanmigo detects a flaw in a student's argument, the parallel AI might generate an instruction to ask the student to reflect on his or her argument to find the flaws. When tutoring in algebra, it undergoes a sophisticated process to ensure that its own solution to a problem is correct. The parallel AI also issues an instruction not to tell a student that he or she has made a mistake, but instead, to ask them to work carefully through, and explain, the process that led to the incorrect solution.

Khanmigo does not produce writing for students. Instead, it supports them develop their own writing skills. It can identify grammatical errors, critique arguments, and draw attention to claims that appear poorly supported by the evidence the student supplied. It also provides support for reading comprehension by asking students questions about a text to check understanding. Again, rather than training students to rely on technology, Khanmigo supports their engagement in deep learning. Khanmigo can also motivate students to write, by coauthoring with them. For example, it might ask a student if he or she would like to collaborate to write a story.

It suggests options for a genre, and then asks the student to begin the story with two sentences. The AI then follows with another two sentences, taking turns with the student.

In the TED talk, Khan described an example of Khanmigo, almost literally, bringing a great novel to life for a student. The student told the AI to adopt the persona of Jay Gatsby, the protagonist of F. Scott Fitzgerald's novel *The Great Gatsby*. The student then asked 'Gatsby' why he kept staring at a green light across the water, as described in the novel. Khanmigo, in the persona of Jay Gatsby, responded, "It's situated at the end of Daisy Buchanan's dock … it represents my yearning for the past and my hope to reunite with Daisy …". It then asked the student for a reflection on a dream or desire "that just seemed out of reach".[54] According to Khan, after a prolonged conversation, the student thanked 'Gatsby' and apologised for taking his time – a testament to the engagement the tool had evoked.

Knowing that Khanmigo is an AI, students have proven willing to engage with it in ways that they might hesitate to engage with teachers or peers, for fear of judgement. Khanmigo can debate with students, giving them opportunities to test their controversial ideas. Given the sensitivity to causing offence felt by many young people in recent years, this seems a valuable feature of the platform.[55] There is a risk here, however, which Lukianoff and Haidt's work highlights.[56]

If AI enables students to bypass learning how to negotiate the harsher realities of human interaction, they may not develop the social resilience needed for adult success.

There are many potential motivational benefits to AI platforms like Khanmigo. Khan explained that Khanmigo can answer questions like, "why do I need to learn this?"[57] using an example where a student asked why they needed to learn about the sizes of different types of biological cells. Khanmigo responded by asking about the student's aspirations. When the student expressed an interest in professional athletics, Khanmigo related the application of cell biology to nutrition, exercise and other physiological processes relevant to athletics.

Khanmigo has distinct ways of interacting compared to more generic AI platforms, such as ChatGPT. It is a specialised educational tool developed by education experts cognisant of the mnemonic, attentional and motivational characteristics of human learning. Its design illustrates the importance of customising AI platforms for educational purposes, rather than allowing generic platforms to dominate. According to Khan, good educators must advocate for beneficial educational AI applications, and develop well-designed approaches for delivery. He quipped that they must use AI to enhance, not undermine, 'HI' – human intelligence.

# CHAPTER 5
# AI as a support for teaching

Teachers require a wide range of knowledge and skills. They must have strong subject knowledge and be able to effectively convey that knowledge to students. They must be able to collect and interpret information on each student's learning and progress to identify learning gaps and weaknesses and know how to correct them. They must offer students specific, timely feedback. Professor John Hattie's 2009 second-order meta-analysis showed that such feedback is the most powerful pedagogical tool at teachers' disposal.[58] Teachers must also be able to form relationships of trust with their students as well as maintain orderly, convivial classrooms.

Generative AI can contribute to some, but not all, of these functions. Educative relationships and communities cannot be replaced by technology. Experience during and following the COVID lockdowns has demonstrated this in tragic ways. Without the same level of relationship and community that is possible in physical classrooms, many students have disengaged from school. New Zealand is now in the grip of a truancy crisis. Less than half of enrolled students attended school regularly in Term 3, 2023.[59] As a result, the informal social education that comes with school attendance has been damaged. Although truancy in New Zealand had been increasing for several years prior to the pandemic, the closure of schools during lockdowns exacerbated the problem and reversing it is proving difficult.

AI will not make human teachers redundant. Teachers will always need a strong command of the material they teach, the skills to teach it effectively, and the ability to build strong relationships with their students. It can, however, assist teachers in their work. In this section, three areas in which AI promises to contribute to the work of teachers are discussed. These are the provision of formative feedback, the collection and analysis of student achievement data, and credentialling of learning.

## 5.1 Formative feedback

Khanmigo is described by Sal Khan as a personal tutor, designed to function with the oversight of a human teacher. An especially powerful function of AI tools like Khanmigo is the provision of formative feedback.

AI can produce writing that is better structured at all levels – sentence, paragraph and whole-document – than anything of which most school students are capable. A risk is that students use it in a similar way to that in which many have used the internet for some time: Faced with an information-based writing task, many students are inclined to copy and paste material from web pages, often unedited.

In the era of large-language models, rather than copying and pasting information for, say, a history essay about the origins of WWII, now they will be tempted simply to ask a chatbot to write it for them. The result would be well composed, (mostly) factually correct, flawless in its grammar and syntax – and nothing substantial would have been learned.

Instead, consider the following scenario:

> A student consults various online sources, some high-quality, others less so, and composes an essay based on these sources. Some of it is well written. These are largely the parts that

the student has copied and pasted from the internet. Other parts, those written by the student, are not so well written. There are problems with both grammar and composition. The student has expressed a few good ideas, although they could be better developed.

The student then submits the draft essay to an AI tutor, which provides feedback on many aspects of the essay. It highlights material that has been copied from the internet, with a warning about plagiarism. It invites the student to learn about paraphrasing, which it models, using some of the plagiarised material. The student then paraphrases the remaining plagiarised material, with the AI tutor providing formative feedback on the fly.

Next, the AI tutor offers advice on composition. It does not simply restructure the essay, although it could. Instead, it describes the elements of a well-structured essay, focusing on the aspects that are poor in the draft essay, and provides some general advice for improving it. The student then sets about restructuring the document, with ongoing formative feedback from the AI.

The AI tutor analyses the essay for its predominant problems in sentence structure. Perhaps the sentences are too long, contain mixed tense, or contain too many colloquialisms. It then highlights the main issues and demonstrates how to fix a few examples of each. The AI highlights any references of dubious quality and suggests better-quality alternatives. It points out where a conclusion may not be adequately supported by an argument and offers alternative interpretations of some of the evidence.

This kind of AI tool would not be doing the student's work; instead, it would be providing the feedback the student requires to become a better researcher and writer. Even so, tools like this must have oversight from a skilled teacher. Teachers should treat AI as an assistant rather than a colleague and maintain relationships with all their students. The human dimension of pedagogy must not be lost.

A sound pedagogical relationship entails intellectual, emotional and social dimensions. Arguably, AI can fulfil the first of these; it cannot fulfil the second or third. There is already a risk of students coming to rely on AI to substitute for human relationships beyond the classroom. Educational AIs must not be allowed to exacerbate that risk. Students must not to come to believe that an AI cares about them or their learning. One role of teachers is to remind students of that, both explicitly, and implicitly through high-quality human interactions.

## 5.2 Assessment and evaluation

AI can make major contributions to teachers' practice through the collection, analysis, collation and reporting of assessment data. This connects to the process of formative feedback: The collection of quantitative and qualitative data, formal and informal, on students' current learning is an essential part of the feedback process.

An especially powerful possibility is that AI could be employed to leverage the statistical power of large data sets and sophisticated analysis to benefit the learning of individual students. There are both technical and political barriers to this, however.

On the technical side, not many teachers are skilled in assessment, and especially not in the analysis of assessment data. Yet, it is a source of valuable information to support learning. If AI can perform technical assessment and analysis functions and present teachers with information that is intelligible to them, and that they can readily use in the classroom, it will make a powerful difference to teaching and learning.

Politically, large-scale, centralised collection of assessment data is viewed with suspicion by many in the teaching profession, especially teachers' unions. They fear it will be used to drive 'performance pay' – promoting, or otherwise financially rewarding teachers, based on the assessment results of their students. Whether or not this fear is well-founded, and irrespective of the merits or otherwise of performance pay, AI could insulate large scale assessment data from that kind of use and leverage its power as a formative tool for individual teachers and students.

AI engines could collect, analyse and collate large datasets of students' work across many schools. This could include both formal assessment data and in-class projects, essays and exercises. An appropriately configured AI would have the capacity to perform comparative judgements on randomly selected pairs of exemplars to calibrate scales measuring a wide variety of skills.[60]

Comparative judgment involves deciding which of a pair of items is superior on a selected dimension. For example, essays could be separately compared on the quality of the writing and on the sophistication of the ideas expressed. If enough such judgments are made on related items, the resulting data can be used to calibrate a measurement scale using the Law of Comparative Judgment.[61] These scales have equal-interval properties: An interval of a given numerical magnitude represents the same amount of educational progress anywhere on the scale, in the same manner as scales measuring physical properties such as distance and weight. This makes them suitable for measuring student progress. Raw test scores and percentages generally lack this equal-interval property.

Having calibrated a scale using a large data set collected across many schools, an AI could then locate individual students on that scale, by comparing new work they submit with stored exemplars. The scales can be used to compute progress rates for individual students, and average rates for whole classes and year levels. Because different scales can be calibrated to measure different aspects of learning, they could be used to identify relative strengths and weaknesses in the learning of individual students. Teachers could use this information to evaluate how effectively they are teaching specific skills and concepts. Similarly, schools could use AI to monitor the quality of their teaching. An AI could collate these things into plain-language reports with simple data displays. The language capability of generative AI could produce descriptive reports for teachers as a supplement to the quantitative analyses. This would be especially helpful to teachers in non-numerate disciplines, some of whom struggle with data displays.

AI engines could create profiles of students across many learning domains. The measurement scales it has calibrated would each represent an axis in a high-dimensional space. By clustering students with similar profiles in this space, AI engines trained on a corpus of valid pedagogical research literature could offer tailored strategies to teachers for addressing the learning needs of individual students.

AI engines could also generate formative feedback for teachers on their practice. Generative AI tools could ask teachers questions, in text or verbally, about actions they have taken to address students' specific learning needs. Informed by data monitoring those students' progress and the pedagogical research literature, the tools could offer more refined advice. If a particular strategy appeared not to be working, they could offer alternatives. As Sal Khan has suggested, AI could also assist teachers with lesson plans and progress reports, freeing their time for human interaction with students.

In addition to applications for formative assessment, AI can also support summative usage of assessment; that is, assessment used to report the outcomes of courses of study. As noted in the opening section of this report, the implications of AI for the authenticity of summative assessment have generated much public debate.

A primary consideration for assessment is its validity. The concept of assessment validity has a vast literature devoted to it. A working definition is that valid assessment enables correct inferences about a student's learning on the dimension of performance it is intended to measure.

A principle for the validity of AI as a summative assessment tool is that anything related to the assessment construct must be independently demonstrated by an assessment candidate. For example, if writing ability is an assessment construct, AI should not be used to assist students to craft written responses. Conversely, if writing is not a test construct, then difficulty with writing may constitute an invalid barrier to demonstrating proficiency on the target construct. A physics examination, for example, may be focused on physics concepts yet require written responses. In this case, it may be appropriate to enable poor writers to demonstrate their knowledge of physics by dictating spoken responses and using voice-to-text and AI writing support.

AI could contribute to summative assessment by enhancing computer adaptive testing (CAT). CAT works by selecting test items from calibrated banks of items based on a candidate's history of responses in the same test. The CAT selects each item at a level of difficulty commensurate with a candidate's ability, estimated on the basis of performance on previous items. Because selection depends on previous performance, CATs determine whether responses are correct or incorrect in real time. They have, therefore, traditionally used a multiple-choice format, which facilitates straightforward determination of correct and incorrect responses. If generative AI could reliably rate students' written responses, it would enable a much wider variety of formats in CATs.

## 5.3 Credentialling

An interesting possibility suggested by Kevin Bell and Dave Moskowitz in the EdTech webinar (see 4.1) is the potential for AI to obviate a need for formal testing. The comparative judgement process described above provides a possible mechanism. Interactions with an AI tutor (like Khanmigo) could be transformed into scale scores on a wide variety of dimensions. If these scales were shown to have sufficient reliability and construct validity, they could credential students who demonstrate proficiency at a criterion level on the scale. The student would not need to complete any formal assessment. In this way, large language AI coupled with psychometric analysis, like comparative judgement, could seamlessly integrate formative assessment, summative assessment and credentialling. This approach, however, would have to be carefully developed to ensure its reliability and validity.

A less direct application of AI in education is at the systems level. Data can be gathered from education providers and supplemented with other relevant data, especially socioeconomic variables. The ability of AI networks to find higher-order correlations and patterns amongst such variables is likely to yield insights that traditional statistical analysis may not. New Zealand has the benefit of the internationally unique Integrated Database Infrastructure (IDI) maintained by Stats New Zealand.[62] The IDI connects data held by every government agency at the level of individual New Zealanders. The potential power of AI-driven analysis incorporating these data is immense. Such systems level analysis can be used for early identification of children, or schools, at risk, and to target resources and intervention accordingly. At the least, it could provide suggestive insights for further investigation and identify factors that might otherwise be overlooked.

## CHAPTER 6

# A concluding principle

An important principle for the application of any technology to education – including AI – is to ensure that students have mastered, to the point of cognitive automaticity, any knowledge and skills prerequisite to later learning.

This principle is at odds with the paradigm of '21st-century learning' that underlies much of the current discussion on the potential role of AI in education. While commentators have not explicitly argued that students no longer need to learn numeracy, literacy or disciplinary knowledge, many have downplayed their importance. These commentators typically argue that AI can free students to create content and think critically without getting bogged down in difficult learning. The changing nature of work, including the increasing importance of 'soft skills', is often cited to justify deemphasising disciplinary learning. However, this argument relies on the false notion that creativity and criticality are possible without knowledge.

The valid role of AI in education is the same as that of any technology: It should be used to enhance, rather than replace, disciplinary learning. Khan Academy's Khanmigo is an example. It is based on scientific learning principles, so as to enhance, rather than undermine, disciplinary learning.

In addition to highlighting the risks AI poses to education, and some utopian ideas for its use, this report has explored ways in which AI can support the work of skilled teachers. Ultimately, the educative process involves human relationships. AI should not be permitted to usurp the role of the knowledgeable teacher, skilled in imparting knowledge to young people.

# Endnotes


1 For example, J. Oravec. "Artificial Intelligence Implications for Academic Cheating: Expanding the Dimensions of Responsible Human-AI Collaboration with ChatGPT and Bard", *Journal of Interactive Learning Research 34(2)* (2023), 213–237.

2 L. Mearian (2023, April 25). "Schools look to ban ChatGPT, students use it anyway" *Computer World.* Website.

3 For example, J. Hattie. *Visible learning: A synthesis of over 800 meta-analyses related to achievement.* (London: Routledge, 2009).

4 For example, J. Sweller, J.J.G. van Merriënboer & F. Paas. "Cognitive Architecture and Instructional Design: 20 Years Later," *Educational Psychology Review 21(2)* (2019), 261–292.

5 For discussion see, J. S. Bowers. "Parallel Distributed Processing Theory in the Age of Deep Networks", *Trends in Cognitive Sciences 21 (12)* (2017), 950–961.

6 See, Jerry A Fodor. *Modularity of Mind: An Essay on Faculty Psychology.* (Cambridge, Massachusetts: MIT Press, 1983).

7 See, J.R Stroop. "Studies of interference in serial verbal reactions", *Journal of Experimental Psychology. 18(6)* (1935), 643–662.

8 For example, D.E. Rumelhart & J.L. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* (MIT Press. 1986).

9 For example, M.S Seidenberg, M. S.. "A connectionist modeling approach to word recognition and dyslexia", *Psychological Science, 4,* (1993), 299–304.

10 For example, A. Cleeremans & J.L. McClelland. "Learning the Structure of Event Sequences", *Journal of Experimental Psychology: General, 120* (1991), 235–253.

11 Data Science (2023, Feb 3). "Specifics about ChatGPT's Architecture", *StackExchange Data Science.* Website.

12 N. Chomsky. *Knowledge of Language: Its Nature, Origin and Use.* (New York: Praeger, 1986).

13 For example, N Evans & S.C. Levinson. "The myth of language universals: Language diversity and its importance for cognitive science", *Behavioral and Brain Sciences 32(5)* (2009), 429–48.

14 N. Chomsky, I. Roberts & J. Watumull (2023, March 12). "The false promise of ChatGPT". *The New Zealand Hearld.* Website.

15 H. L. Dreyfus. *Alchemy and Artificial Intelligence.* (Rand Corporation, 1965); R. Fjelland. "Why general artificial intelligence will not be realized", *Humanities and Social Sciences Communications 7(10),* 2020.

16 Some animals, such as Bonobos, have shown limited capacity to communicate symbolically after extensive training.

17 See, J.J. Gibson. *The ecological approach to visual perception.* (Houghton, Mifflin and Company, 1979).

18 Ibid. 127.

19 For example, D.C. Dennett. "Reprint of Intentional systems in cognitive ethology: The Panglossian paradigm defended", *The Brain and Behavioral Sciences 6 (3)* (1987), 343–390.

20 For a review see, Banks, S.A. *A Historical Analysis of Attitudes Toward the Use of Calculators in Junior High School.* Master of Education dissertation (Cedarville University, Ohio, 2011).

21 For example, Mogari, D. & S. Faleye. "Introducing calculators to learners early in their schooling: The effect on long-term arithmetic proficiency", *African Journal of Research in Mathematics, Science and Technology Education, 16(3)* (2012), 363–375.

22 Pribram, K.H., G.A. Miller. & E. Galanter. *Plans and the structure of behavior.* (New York: Holt, Rinehart and Winston, 1960); for example, Baddeley, A .D. & G. Hitch. "Working memory", in G.H. Bower (ed.), *The psychology of learning and motivation: Advances in research and theory* (Volume 8) (New York: Academic Press, 1974), 47–89.

23 R. C. Atkinson & R. M. Shiffrin. "Human memory: A proposed system and its control processes", in K. W. Spence, & J. T. Spence (eds.), *The psychology of learning and motivation* (Volume 2) (New York: Academic Press, 1968), 89–195.

24 Cowan, N. "The magical number 4 in short-term memory: a reconsideration of mental storage capacity", *The Behavioral and Brain Sciences 24 (1)* (2001), 87–185.

25 J. Sweller, P. Ayres & S Kalyuga. *Cognitive Load Theory*. (Springer, 2011). 4.

26 See, J. Sweller (2016). "Cognitive load theory, evolutionary educational psychology, and instructional design", in D. Geary & D. Berch (eds). Eds. *Evolutionary perspectives on child development and education.* (Basel: Springer), 291–306.

27 For example, N. Serki & S San Bolkan. "The effect of clarity on learning: impacting motivation through cognitive load," *Communication Education,* (2023), 1–17.

28 D.C. Geary. *The Origin of the Mind: Evolution of Brain Cognition and General Intelligence.* (American Psychological Association, 2004).

29 EdTech NZ (Host) (2022, February 22). *Artificial Intelligence and the Impact on Education.* Online EdTech NZ and NZTech Connect Event. Website.

30 Khan Academy. Supercharge your teaching experience with Khanmigo. Website (2024).

31 TTRO Never Stop Learning. (2024). Website.

32 EdTech NZ (Host) (2022, February 22). *Artificial Intelligence and the Impact on Education.* op cit. 13:42mins.

33 Moskowitz, D. (no date). Website.

34 EdTech NZ (Host) (2022, February 22). *Artificial Intelligence and the Impact on Education.* op cit. 50:20mins.

35 Ibid. 51:05 mins.

36 Ibid. 50:36mins.

37 J. Sweller et al. *Cognitive Load Theory.* op. cit. 32.

38 The Mind Lab. 2024. Website.

39 EdTech NZ (Host) (2022, February 22). *Artificial Intelligence and the Impact on Education.* op cit. 29:00 mins.

40 AI Forum New Zealand. 2022. Website.

41 S. May & E. Medina. *PISA 2022 Aotearoa New Zealand Summary Report.* (Ministry of Education, 2023).

42 See, R. Patterson. *Un(ac)countable: Why millions on maths returned so little.* (The New Zealand Initiative, 2015).

43 For example, Mogari, D. & S. Faleye. "Introducing calculators to learners early in their schooling: The effect on long-term arithmetic proficiency". op.cit.

44 Jordan, N.C., D. Kaplan, C. Ramineni & M.N. Locuniak. "Early math matters: Kindergarten number competence and later mathematics outcomes", *Developmental Psychology 45(3)* (2009), 850–867.

45 EdTech NZ (Host) (2022, February 22). *Artificial Intelligence and the Impact on Education*. op cit. 27:05 mins.

46 Ibid. 16:26 mins.

47 Ibid. 17:19 mins.

48 Ibid. 53:00 mins.

49 Ibid. 18:19 mins.

50 Ibid. 32:37 mins.

51 B. Bloom. "The Two-Sigma Problem: The Search for Methods of Instruction as Effective as One-to-One Tutoring", *Educational Researcher 13(6)* (1984), 4-16.

52 Khan Academy. "Supercharge your teaching experience with Khanmigo". op cit.

53 S. Khan. (Speaker) (2023, April). *How AI could save (not destroy) education.* TED Conferences LLC. Website.

54 Ibid. 6:20-6:51 mins.

55 See for example, G. Lukianoff & J. Haidt. *The Coddling of the American Mind.* (Penguin Press, 2018).

56 G. Lukianoff & J. Haidt. *The Coddling of the American Mind.* op. cit.

57 S. Khan. (Speaker) (2023, April). *How AI could save (not destroy) education.* op. cit 4:33 mins.

58 J.Hattie. *Visible learning: A synthesis of over 800 meta-analyses related to achievement.* op. cit.

59 Education Counts. *Attendence* (2023, December). Website.

60 See for example, Tarricone, P., & C.P. Newhouse. "Using comparative judgement and online technologies in the assessment and measurement of creative performance and capability", *International Journal of Educational Technology in Higher Education, 13(16)* (2016).

61 L.L. Thurstone. "A law of comparative judgement", *Psychological Review 34* (1927), 273-286.

62 Stats NZ Tatauranga Aotearoa. *Integrated Data Infrastructure.* (New Zealand Government, 23 August 2022). Website.

# Bibliography


AI Forum New Zealand. 2022. Website. https://aiforum.org.nz/about/our-people/

Atkinson, R. C., & R. M Shiffrin. "Human memory: A proposed system and its control processes", in K. W. Spence, & J. T. Spence (eds.), *The psychology of learning and motivation* (Volume 2) (New York: Academic Press, 1968), 89–195.

Baddeley, A .D. & G. Hitch. "Working memory", in G.H. Bower (ed.), *The psychology of learning and motivation: Advances in research and theory* (Volume 8) (New York: Academic Press, 1974), 47–89.

Banks, S.A. *A Historical Analysis of Attitudes Toward the Use of Calculators in Junior High School.* Master of Education dissertation (Cedarville University, Ohio, 2011). https://digitalcommons.cedarville.edu/cgi/viewcontent.cgi?article=1030&context=education_theses

Bloom, B. "The Two-Sigma Problem: The Search for Methods of Instruction as Effective as One-to-One Tutoring", *Educational Researcher 13(6)* (1984), 4–16.

Bowers, J.S. "Parallel Distributed Processing Theory in the Age of Deep Networks", *Trends in Cognitive Sciences 21 (12)* (2017), 950-961. ISSN 1364-6613.

Chomsky, N. *Knowledge of Language: Its Nature, Origin and Use*. (New York: Praeger, 1986).

Chomsky, N., I. Roberts, & J Watumull. (2023, March 12). "The false promise of ChatGPT". *The New Zealand Hearld*. Web. 8 April 2014. https://www.nzherald.co.nz/world/the-false-promise-of-chatgpt/YZLSLE5HXBAQVCWLLIOPC73KXU/

Cleeremans, A. & J. L. McClelland. "Learning the Structure of Event Sequences", *Journal of Experimental Psychology: General 120* (1991), 235-253. http://dx.doi.org/10.1037/0096-3445.120.3.235

Cowan, N. "The magical number 4 in short-term memory: a reconsideration of mental storage capacity", *The Behavioral and Brain Sciences 24 (1)* (2001), 87–185. DOI:10.1017/S0140525X01003922. PMID 11515286.

Data Science (2023, Feb 3). "Specifics about ChatGPT's Architecture", *StackExchange Data Science.* Web. 9 April 2024. https://datascience.stackexchange.com/questions/118273/specifics-about-chatgpts-architecture

Dennett, D. C. "Reprint of Intentional systems in cognitive ethology: The Panglossian paradigm defended", *The Brain and Behavioral Sciences 6 (3)* (1987), 343–390. DOI:10.1017/s0140525x00016393. S2CID 32108464.

Dreyfus, H. L. *Alchemy and Artificial Intelligence.* (Rand Corporation, 1965). https://www.rand.org/content/dam/rand/pubs/papers/2006/P3244.pdf

EdTech NZ (Host) (2022, February 22). *Artificial Intelligence and the Impact on Education.* Online EdTech NZ and NZTech Connect Event. https://edtechnz.org.nz/event/artificial-intelligence-and-the-impact-on-education/

Education Counts. *Attendence*. (2023, December). https://www.educationcounts.govt.nz/statistics/attendance

Evans, N. & S.C. Levinson. "The myth of language universals: Language diversity and its importance for cognitive science", *Behavioral and Brain Sciences 32(5)* (2009), 429–48.

Fjelland, R. "Why general artificial intelligence will not be realized", *Humanities and Social Sciences Communications 7(10)*, 2020. https://doi.org/10.1057/s41599-020-0494-4

Fodor, Jerry A. *Modularity of Mind: An Essay on Faculty Psychology.* (Cambridge, Massachusetts: MIT Press, 1983). ISBN 0-262-56025-9

Geary, D.C. *The Origin of the Mind: Evolution of Brain Cognition and General Intelligence.* (American Psychological Association, 2004). ISBN: 978-1591471813

Gibson, J. J. *The ecological approach to visual perception.* (Houghton, Mifflin and Company, 1979).

Hattie, J. "Visible learning: A synthesis of over 800 meta-analyses related to achievement" (London: Routledge, 2009).

Jordan, N. C., D. Kaplan, C. Ramineni & M. N. Locuniak. "Early math matters: Kindergarten number competence and later mathematics outcomes", *Developmental Psychology 45(3)* (2009), 850–867. https://doi.org/10.1037/a0014939

Khan Academy. "Supercharge your teaching experience with Khanmigo". Website (2024). https://www.khanacademy.org/khan-labs

Khan, S. (Speaker) (2023, April). *How AI could save (not destroy) education*. Ted Conferences LLC. Website. https://www.ted.com/talks/sal_khan_how_ai_could_save_not_destroy_education

Lukianoff, G. & J. Haidt, J. *The Coddling of the American Mind*. (Penguin Press, 2018).

May, S. & E. Medina. *PISA 2022 Aotearoa New Zealand Summary Report*. (Ministry of Education, 2023). https://www.educationcounts.govt.nz/__data/assets/pdf_file/0015/224601/PISA-2022-summary-report.pdf

Mearian, L (2023, April 25) "Schools look to ban ChatGPT, students use it anyway" *Computer World.* Website. https://www.computerworld.com/article/3694195/schools-look-to-ban-chatgpt-students-use-it-anyway.html

Mogari, D. & S. Faleye. "Introducing calculators to learners early in their schooling: The effect on long-term arithmetic proficiency", *African Journal of Research in Mathematics, Science and Technology Education 16(3)* (2012), 363-375. DOI: 10.1080/10288457.2012.10740751

Moskowitz, D. (no date). Website. https://dave.moskovitz.co.nz/?s=students

Oravec, J. "Artificial Intelligence Implications for Academic Cheating: Expanding the Dimensions of Responsible Human-AI Collaboration with ChatGPT and Bard", *Journal of Interactive Learning Research 34(2*) (2023), 213-237. https://philarchive.org/archive/ORAAII

Patterson, R. *Un(ac)countable: Why millions on maths returned so little.* (The New Zealand Initiative, 2015). https://www.nzinitiative.org.nz/reports-and-media/reports/unaccountable-why-millions-on-maths-returned-little/document/20

Pribram, K.H., G.A. Miller & E. Galanter. *Plans and the structure of behavior*. (New York: Holt, Rinehart and Winston, 1960). ISBN 978-0-03-010075-8. OCLC 190675.

Rumelhart, D.E. & J. L. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* (MIT Press, 1986). ISBN: 9780262680530

Seidenberg, M. S. "A connectionist modeling approach to word recognition and dyslexia", *Psychological Science* 4, (1993), 299–304.

Serki, N. & S. San Bolkan. "The effect of clarity on learning: impacting motivation through cognitive load," *Communication Education* (2023), 1-17. DOI: 10.1080/03634523.2023.2250883

Stats NZ Tatauranga Aotearoa. *Integrated Data Infrastructure*. (New Zealand Government, 23 August 2022). Website. https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/

Stroop, J. R. "Studies of interference in serial verbal reactions", *Journal of Experimental Psychology. 18(6)* (1935), 643–662. DOI:10.1037/h0054651.

Sweller, J., P. Ayres & S. Kalyuga. *Cognitive Load Theory.* (Springer, 2011). DOI: 10.1007/978-1-4419-8126-4

Sweller, J (2016). "Cognitive load theory, evolutionary educational psychology, and instructional design", in D. Geary & D. Berch (eds.), *Evolutionary perspectives on child development and education* (Basel: Springer), 291–306.

Sweller, J., J.J.G. van Merriënboer & F. Paas. "Cognitive Architecture and Instructional Design: 20 Years Later," *Educational Psychology Review 21(2)* (2019), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tarricone, P., & C.P. Newhouse. "Using comparative judgement and online technologies in the assessment and measurement of creative performance and capability", *International Journal of Educational Technology in Higher Education 13(16)* (2016). https://doi.org/10.1186/s41239-016-0018-x

The Mind Lab. 2024. Website. https://academyex.com/faculties/themindlab

Thurstone, L.L. "A law of comparative judgement", *Psychological Review 34* (1927), 273-286.

TTRO Never Stop Learning. (2024). Website. https://www.ttro.com/

Artificial Intelligence (AI) is impacting many aspects of society and the economy, including education. If it is used appropriately, it will increase teachers' productivity and contribute to improvements in students' learning. If not, it will degrade students' learning and teachers' practice.

This report presents a framework, based on scientific research in human cognition, for the application of AI in the classroom. It challenges the notion, prevalent in discussion about AI in education, that AI will change what is important for students to learn and the way in which they learn.

AI can make important contributions to formative feedback, assessment and evaluation and, potentially, credentialling of learning. However, it must not be allowed to usurp the role of the knowledgeable teacher, skilled in imparting knowledge to young people.

A general principle for the adoption of AI and other technology in education is that AI must not be a substitute for learning knowledge and skills prerequisite to later learning. Such learning must be securely encoded in long-term memory.

**$25.00**

**ISBN**
978-1-7386277-0-7 (print)
978-1-7386277-1-4 (online)

THE NEW ZEALAND INITIATIVE

www.nzinitiative.org.nz

**The New Zealand Initiative**
PO Box 10147
Wellington 6143