Submission to The House Standing Committee on Employment, Education and Training Inquiry into the use of generative artificial intelligence in the Australian education system.

Associate Professor Erica Southgate, School of Education, University of Newcastle, Australia.

1. The proliferation of user-facing and system-powering artificial intelligence (AI) applications is and will continue to challenge fundamental norms and expectations that are foundational to Western education systems from early childhood to higher education. Educational institutions use AI powered applications that are generalised (for example teleconferencing platforms, design and media production applications) and that are specific to educational technology such as learning management systems, adaptive tutoring and assessment applications, online proctoring, pedagogical agents embedded into platform to assist users, and learning analytics dashboards shaped by algorithms. In addition, AI business intelligence and other machine learning is being used behind the scenes by education system and institution administrators and policy makers to inform decisions which most stakeholders are not always privy to. It is difficult for ordinary humans such as students and their families, educators and student services facilitators - all of whom allow schools, VET and higher education institutions to run in fair and optimal ways - to gain information from those in executive and policy making roles on the use of AI systems in decision making and representations of learning or job performance. Moreover, this lack of transparency is coupled with opaque or inadequate accountability mechanisms – who is responsible for ensuring AI systems do not produce biased or discriminatory outcomes, and if they do, how with students, families and staff know if it has occurred and where to seek rectification of harm? This cuts across all levels of education in Australia. There is much policy, regulatory and industry standards work to be undertaken on AI in education in general. Only focusing on the effects of generative AI in education is myopic. It is like being dazzled by one fascinating tree in a large and often impenetrable forest. In the context of managing an ecosystem that can have a profound impact on humans, especially vulnerable humans, I would urge the Inquiry to acknowledge the extent and significance of the AI ecosystem in education, its state-of-play, the asymmetrical power relations between stakeholders, as well as the potential and risks for learning and learners and teachers, so that the Australian people are served by a forward thinking, big picture response that grapples with all the complexity, not just one type of AI.

2. Generative AI is only the first of what will be many disruptive use cases for education. Educators are going through the same sense of disorientation that journalists first went through more than five years ago with the advent of bot-powered disinformation and deep fakes on the internet. The current sense of disorientation and grasping for policy-positions on generative AI was not an inevitable consequence of unpredictable innovation. It should not have taken education policy makers or administrators by surprise. We might not have known the exact timeframe for the large-scale commercial release of generative models, but we did know it was coming. We knew because: the field of AI in ED has been around since the 1970s; big technology companies regularly release research papers and blog posts and lodge patents that provide a window into the direction of their AI innovation including generative AI; there were a number of excellent researchers including those in the Australian context who have been writing and expert commentating in the field for years (Neil Selwyn, Kal Gulson, Carlo Perrotta, Jane Hunter, Maarten de Laat, Matt Bower); there was a major report commissioned by the Australian Government Department of Education in 2018 which included a robust ethics and governance framework for AI in education called *"Artificial Intelligence and Emerging Technologies (virtual, augmented and mixed reality) in Schools Research Report"* (Southgate et al., 2018); and a wide ranging discussion paper on the implications of AI for equity and higher education was released in 2020 (Southgate, 2020). In 2019, I published "AI did my homework" (Southgate, 2019a) and "Educators! it's time to talk about how artificial intelligence will rock our

world" (Southgate 2019b) in popular education forums and in 2021 I published '*Deep fakes, authenticity and authentication in the age of artificial intelligence*" in a well-known teacher's union magazine (Southgate, 2021). More recently I co-authored an article with innovative NSW English teacher Imelda Judge on AI for teachers which looked specifically at generative AI (Southgate & Judge, 2022). Over the years my voice was one amongst others in the Australian context, advocating for an urgent, sustained, and collaborative approach to preparing schools, VET and universities for the challenges and opportunities of AI for education. This included the urgent need for serious examination of the challenges of generative AI to the foundational concept of Western education systems that originality and authenticity are the measure of quality and competency for student work and assessment. It also included the need to closely examine how the adaptive personalisation of AI-power application might benefit students with special needs – an area that receives generally very scant attention.

3. Leaving aside those that have decided that generative AI should be banned in schools, but who seem less vocal on students using it for homework, there is currently several (not mutually exclusive) positions on generative AI evident amongst educators. These are:

~ *Generative AI is the cool new tool for school*: Educators in this camp have undertaken some excellent work in developing curriculum-aligned ideas for using generative AI in classrooms, especially in schools where it is not banned. This position is generally enthusiastic and focused on the practicalities of helping students understand what generative tools can do well and where they are fraught, and how they can be creatively and critically used to augment learning. There is certainly scope within the ICT, Ethical Understanding, and Critical and Creativity Thinking Capabilities frameworks of the Australian Curriculum to engage with such tools, although we really don't have any idea to what extent and quality the Capabilities frameworks are implemented in schools. This position is optimistic stems from the idea of technology as learning tool.

~ *Access to AI tools is a matter of equity*: This position, closely related to that above, highlights the need for equitable access to new technologies for learning and highlights the potential for a new digital divide to open-up between those students allowed to develop capabilities using AI tools and those banned from doing so. It is heartbreaking that the area of the digital divide in Australian schooling (what new hardware, software, expertise and professional learning related to technology enhanced learning) generally receives very attention from policy makers and researchers. If you visit a range of schools across the public and private sectors the educational technology divide, especially in the emerging technology divide in terms of student access to extended reality (VR and AR), robotic and AI products and the school building infrastructure to use these for learning in optimal ways, is truly disturbing. It is time to seriously attend to the digital divide in Australian schooling in terms of funding, policy and research.

~ *AI-proofing assessment is imperative*: This position, especially in senior schooling and higher education where plagiarism is of most concern, focuses on trying to AI-proof assessment by assessing the learning process and not just the assignment product, requiring proof of the learning process, revising declarative knowledge assessment to more authentic tasks which require application and transfer of knowledge to context, and supplementing assessment with oral explanatory component to test knowledge and understanding. This position acknowledges that at present plagiarism checking software is not altogether reliable in identifying AI generated work. Educators in this camp have received sometimes confusing advice from administrators on the effectiveness of plagiarism software to detect AI texts and often have been afforded very little time to revise curriculum and assessment to make it more AI resistant. Some in this group realise that students can use a generative AI tool to create a text which is then put through paraphrasing

software (also powered by AI) to produce a written assignment that has a good chance of avoiding detection by plagiarism software. As an aside, I have no doubt that the essay mill business model of cheating is being replaced by an all-in-one product or series of inexpensive compatible apps designed to produce text, music, visual artefacts etc that can beat plagiarism software.

~ *Technological solutionism will not fix the problem*: This position understands that hoping for technological advancements that will save us from a complex problem that has resulted from the intended or unintended consequence of technology is futile without the development and committed implementation of effective design standards, policy, and legislation that can regulate AI and protect human rights. Generative AI models were and are released with minimal care for consequences although technology companies have and continue to highlight the grave risks they pose to social and economic stability, democracy, cyber security, and human dignity (Electronic Privacy Information Centre, 2023). This position acknowledges the absurdism of waiting and relying on the development of a technological solution to the risks of AI because it understands that the history of the market-driven innovation is always to develop a product that can game or beat the problem. Waiting for an AI to adequately detect AI-generated student work fails to understand that another AI-powered product will be developed to get around the detection software and so on and so on.

~ *Forming a position on the ethics of AI and automation in education needs to happen urgently*: The history of AI is littered with examples of bias and discrimination. This can occur for many reasons including lack of diversity in technology workforce to recognise potential bias, lack of diversity in data used to train models, lack of careful trialling and monitoring of products, the "black box' algorithmic complexity of some types of machine learning that prevent auditing or understanding of machine decision-making, and the "black box" of procurement which protects commercial-in-confidence information about algorithms and largely precludes transparent governance of the technology. There are many credible reports that the development of AI including large language models are based on the extreme exploitation of labour in the developing world (Williams, Miceli & Genaro, 2022). It is time to consider if there is a place for educational technology products in our institutions that have been developed through cruel exploitation of humans. In addition, Wachter and colleagues (2020) argue that the automation of fairness through algorithms may not always be possible because AI presents serious challenges in terms of new and perhaps undetectable forms of discrimination. They write:

"Compared to human decision-making, algorithms are not similarly intuitive; they operate at speeds, scale and levels of complexity that defy human understanding, group and act upon classes of people that need not resemble historically protected groups and do so without potential victims ever being aware of the scope and effects of automated decision-making. As a result, individuals may never be aware they have been disadvantaged and thus lack a starting point to raise a claim under non-discrimination law." (Wachter et al., 2020, p.6).

It is vital that we understand that AI systems may discriminate in ways that are without precedent and that there are currently few means of detecting or investigating this to prevent discrimination. Furthermore, Wachter and colleagues argue that this can hinder the collection of evidence to mount a *prima facie* case for new forms of discrimination (for example, that automated discrimination may only be observable at a statistical level, and this may be inaccessible to technical people and non-technical people alike given the proprietary and often opaque nature of AI algorithms and/or the need for highly specialised mathematical knowledge). Hence, we are now entering a world where machines may discriminate in ways that are different to humans, with harm not always discernible in ways humans can conventionally comprehend or that may not be apparent until well after harm has

occurred. AI may very well reshape discrimination as we now understand it and this will could have significant implications for vulnerable populations such as children and young people interacting with AI powered system in school and at home.

4. There has been widespread outcry and organised protest regarding the development of generative AI using indiscriminate harvesting of creative works that violate the moral rights and copyright of creatives and the replacement or exploitation of the creative workforce through AI (the writers and actors strike in the U.S. being a current example). There is little doubt that AI will be radically disruptive to career education generally including the possible decimation of the career pipeline for creative industries. There are more optimistic perspectives where the technology can augment human capabilities and talk of new careers for those that can harness AI. Even the focus on teaching children coding may be largely anachronistic in a world where there are products in which AI codes with humas and a push for more of this work to be done by machines. Career education from schooling to vocational and higher education needs a more intensive focus with more serious resourcing in the AI era if it is to be useful for students.

5. Educational institutions are social places. Learning is a social process. There is research that illustrates that humans anthropomorphise and can overly trust machines. The relationship been student-teacher-machine is being radically transformed by AI as new forms of sociality enter our educational institutions. We are already seeing cases where humans are forming bonds with AIs that go beyond the pragmatic into the affective realm. At a developmental level, children have difficulty discerning reality from fantasy until the ages of around 7-11. User facing AI, especially that designed to mimic human interaction, is being designed to keep us engaged. Teachers will need to make careful decisions about whether this type of AI is suitable for their learners. Some adolescent and adult students will form bonds with AI that may not be good for developing their capability as active, questioning, and agentic learners. Procurement decisions should consider the potential risks of AI applications that covertly 'nudge' students and teachers in particular learning directions and of 'dark pattern' design elements in user interfaces that are crafted to deceive users into taking certain actions. Transparency and explainability should be key here. At its very core teaching is about explaining things – content, pedagogical and curriculum decisions, pastoral choices – and to develop students who can understand and explain what they know and can do and their learning journey. This cherished foundation should not be undermined by automated decision-making. During the Robodebt Inquiry much was made about humans-in-the-loop in automated systems. This phrase is often bandied about when discussing automated decision making. The Robodebt Inquiry demonstrated there were many, many humans-in-the-loop and yet this did not prevent immense harm from occurring. AI won't replace teachers but new relational learning and power triads between student-teacher-machine will be facilitated through the design and implementation of the technology. This design and the promises made by technology companies will focus on establishing and maintaining trust in automated decisioning making. Educators, students, and communities must be empowered through education, policy, and regulation to be able to ask the right questions about the suitability and ethics of a product's design and the robustness and independence of evidence of efficacy for education. They must have ready access to transparent accountability mechanism to contest or decline the use of such technology in classrooms. Even with ring-fenced applications of generative AI tools, where students are using a version of the products restricted to an education system or cohort – transparency and contestability must be key. Students, families, and teachers have the right to ask questions about how exactly privacy is being protected, how student data is being shared and used by the vendor and education system administrators, and what mechanisms are in place to ensure discrimination and bias do not occur. This will be particularly important as the harvesting of biometric data becomes embedded in educational technology applications and

platforms and with the widespread roll-out of smart technology in educational institutions. It is not enough for education policy makers or system administrators to say to students, teachers, and families "Trust us, your privacy and well being is ensured" without the above questions being addressed *in detail and in public*. The risk of cosy relationships with technology companies and the regulatory capture of policy makers and administrators, is very real in the area of AI due to the sheer technical complexity of the technology, the black box nature of many machine learning models and the protection of commercial-in-confidence algorithms. In this space, the ethical issue of regulatory capture and its consequences needs serious attention.

6. There are software applications that can be used to easily manipulate media to generate alternative and fake text, audio, images, and video. AI has and will be deployed in free or cheap applications that can create and distribute deep fakes with relative ease. This is already occurring with deep fake pornography that is used to victimise and ruin the lives of mainly women and girls. Deep fake audio is also being used for malicious and criminal enterprise. Deep fake apps will pose significant challenges to schools and other educational institutions as they are weaponised for bullying, harassment, and deception. The rapid human and bot spread of deep fakes will probably surpass the damage already occurring with student online bullying and will adversely affect staff who are targeted and the ethical culture of the educational institution. The anonymity through which deep fakes can be created will exacerbate the issue.

To conclude, Australia is a signatory to Human Rights conventions including the digital rights of the child. The design, implementation, and governance of AI technology generally and in educational institutions should uphold human rights. This will need substantial new policy, regulation and legislation, design standards and a larger pedagogical project designed to educate everyone about AI and their human rights in relation to it.

Australia has been a signatory to the United Nations Convention on the Rights of the Child (CRC) since 1990. As the most widely ratified human rights treaty in history, the CRC provides 54 articles or principles that set the foundation for allowing all children (those under 18 years of age) to have a healthy and safe childhood free from discrimination and ripe with opportunities for full development. Child rights include the right to an education; to have and share information that is not damaging to them; to have access to reliable information; and privacy. The CRC stipulates that governments should make these rights available to children. In 2021, the United Nations delivered an important document that highlighted the rights of the child in the digital age. It was called *General comment No. 25 on children's rights in relation to the digital environment*. General comment no. 25 provides a broad overview of opportunities and challenges in fulfilling the rights of all children in the digital world, stating:

"The digital environment … [includes] digital networks, content, services, and applications, connected devices and environments, virtual and augmented reality, artificial intelligence, robotics, automated systems, algorithms and data analytics, biometrics and implant technology … [It is] becoming increasingly important across most aspects of children's lives, including during times of crisis, as societal functions, including education, … progressively come to rely upon digital technologies. It affords new opportunities for the realization of children's rights, but also poses the risks of their violation or abuse. During consultations, children expressed the view that the digital environment should support, promote and protect their safe and equitable engagement."

AI powers the technology mentioned in this quote. It is ubiquitous in its integrated into the digital products and platforms children use including at school. Some uses of AI pose low levels of risk, while others are much more problematics. Determining this through a systematic and shared

perspective made public is vital. New types of technologies use sensors, geolocators and artificial intelligence to rapidly harvest information about a person (their demographics, location, and preferences and affiliations) and of a person (bodily attributes and behaviours such as voice, heart rate, finger, limb and eye gaze and movements and even pupil dilation). This information is algorithmically drawn together to create profiles and make predictions about us. This is not just about our online life. It can also impact on the opportunities available to us more generally, as these analytic profiles are used by all types of organisations to automate decisions about what opportunities are, or are not, available to us. The trend is to automate what learning opportunities students have access to and how learners themselves are represented through analytic dashboards. The performance of schoolteachers will also, no doubt, be represented and evaluated through such dashboards — as educators in higher education are now subject to this. This is not necessarily a negative trend if there is regulation, policy and product design that protects our human rights and the rights of the child.

Generative AI is only the first disruptive use case – many more are here now even (if we do not always know where or how machine learning is being used in our educational institutions and by those that administer them) or will arrive shortly. Now is the time for concerted, coordinated, responsible and transparent response by the education sector, one in which students, educators and their families are central to the process not marginalised or told to just trust authorities and the technology companies they deal with. We are fortunate to have good will and deep expertise amongst education researchers and technologists, policy makers and educators for such a response to be developed. This will take a new level of democratic participation and a time frame that extends beyond election cycles but the outcomes for education will be worth it in the end.

References

Electronic Privacy Information Center (2023). Generating harms: Generative AI's Impact and paths forward. Retrieved https://epic.org/wp-content/uploads/2023/05/EPIC-Generative-AI-White-Paper-May2023.pdf

Southgate, E. (2019a) AI did my homework. Retrieved https://ericasouthgateonline.wordpress.com/2019/03/06/ai-did-my-homework/

Southgate, E. (2019b) Educator! It's time to talk about how artificial intelligence will rock our world. *Campus Morning Mail*. Retrieved https://campusmorningmail.com.au/news/educators-its-time-to-talk-about-how-artificial-intelligence-will-rock-our-world/

Southgate, E. (2019c) Artificial intelligence in schools: an ethical storm is brewing. *The Mandarin* Retrieved https://www.themandarin.com.au/author/erica-southgate/ and *EduResearch Matters* and https://www.aare.edu.au/blog/?p=4325

Southgate, E. (2020). Artificial intelligence, ethics, equity and higher education: A 'beginning-of-the-discussion' paper. National Centre in Student Equity and Higher Education (Curtin University). Retrieved http://ncsehe.edu.au/wp-content/uploads/2020/07/Southgate_AI-Equity-Higher-Education_FINAL.pdf

Southgate, E. (2021). Deep fakes, authenticity and authentication in the age of artificial intelligence. Professional Voice 13:3, p. 41-9. Retrieved https://www.aeuvic.asn.au/sites/default/files/Professional%20Voice/PV_13.3/PV_2021_13.3_Complete.pdf#page=41

Southgate, E., Blackmore, K., Pieschl, S., Grimes, S., McGuire, J., & Smithers, K. (2019). Artificial intelligence and emerging technologies (augmented, in schools. Retrieved https://www.education.gov.au/supporting-family-school-community-partnerships-learning/resources/ai-schools-report

Southgate, E. & Judge, I. (2022). The new machine age for English teachers: The pedagogical possibilities and pitfalls of artificial intelligence. *Metaphor*, (2), pp. 10-16.

Williams, A., Miceli, M., & Gebru, T. (2022). The exploited labor behind artificial intelligence. *Noema Magazine*, *13*.

Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review*, *41*, 105567.