COMMENTARY

Constructive Criticism? Working with (Rather than Against) the AIED Back-Lash

Neil Selwyn[1]

Introduction

Artificial intelligence in education (AIED) is currently experiencing a period of increased success *and* heightened scrutiny. On one hand, AI products are now being taken up at scale throughout mainstream schooling and higher education.

A rapidly-growing market for educational AI technologies is being bolstered by venture capital investment, increased policymaker interest, and ongoing enthusiasm that "education will be upended by the data revolution" (Schleicher, 2021, n.p.). In all these ways, triumphant talk of the 'Coming of Age of AIED' might not be too far off the mark.

On the other hand, we are also witnessing growing pushback against the presence of AI technologies in education. The past few years have seen the European Commission designate education to be a 'high risk' area for the application of AI. Various student protests have been staged around the world against personalised learning systems, and teaching unions have campaigned against the classroom imposition of AI-driven technology (e.g. Bowles, 2019, Edelman, 2018, Harwell, 2022). At the same time, a burgeoning academic literature is fuelling dissent and dis-satisfaction around societal applications of AI, including in education.

For some AIED insiders, these recent rebuttals have triggered feelings of annoyance, dismissal and perhaps mild offence. Many others in the AIED community remain nonplussed by these criticisms of their work. Nevertheless, it is well worth engaging with what might initially appear to be annoying, wrong-headed 'misinformation' being peddled by outsiders who lack basic technical knowledge of the technologies they are arguing against. Such controversies and growing critique might well eventually turn out to be nothing more than temporary growing pains. Alternately, these critical arguments might forewarn the demise of the field as we know it. Either way, it seems sensible for everyone in AIED to pay close attention and reflect on what this backlash against AI in education means for future work in the area.

✉ Neil Selwyn
neil.selwyn@monash.edu

[1]   School of Education, Culture & Society, Monash University, Melbourne, Australia

Putting Criticisms of AIED Into Context

First, it is important to note that this growing dissent and disgruntlement is not specific to AI *in education*. Indeed, the past ten years or so has seen rising scepticism around digital technologies in most areas of society – what has become known as the "tech-lash" (Atkinson et al., 2019). There is ongoing controversy around the development of various AI technologies such as facial recognition, natural language generation, lethal autonomous weapons, as well as the application of AI in areas such as Fin-Tech, Ad-Tech and Med-Tech. As such, the 2020s are proving to be a time when research and development in *all* areas of AI is attracting increasing scepticism. As one recent headline put it: "We used to get excited about technology – what happened?" (Vallor, 2022).

Second, it is important to get a clear sense of the distinct interests and agendas that are driving debates around AIED. This is not a case of two opposing sides of those who 'understand' AI versus those who do not. Instead, as a broad guide, we might like to think of at least seven broad sets of different positions at play (see Fig 1):

- **Technical proponents of AIED**: people from computer science, IT and software development backgrounds involved in the technical development of AIED software and systems. This group tends to conceive AI issues in terms of what Birhane et al. (2021) describe as an 'engineering mindset'.
- **Educational proponents of AIED**: people enthused by the potential of AI to support – if not transform - teaching and learning. This group tends to frame
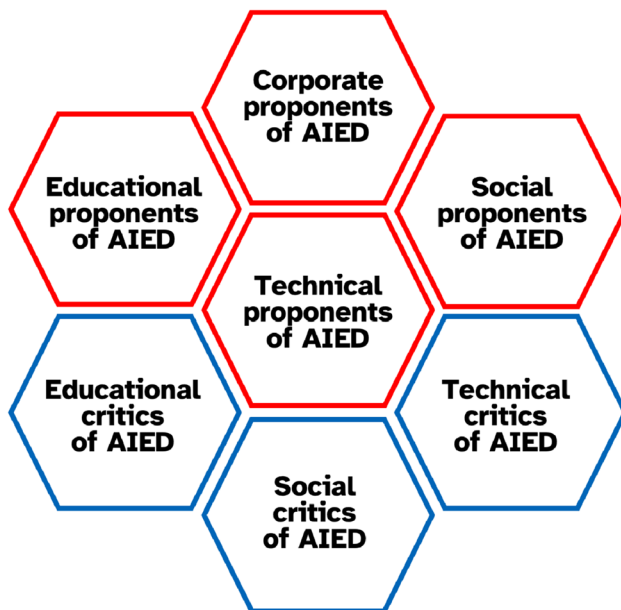

Fig. 1 Who is talking about AIED in the 2020s? A loose matrix of interests and agendas

AIED in terms of various models of learning and teaching, and align with theories from the learning sciences and education psychology.

- **Corporate proponents of AIED**: most obviously, IT industry actors and tech vendors profiting from the sale of AIED products. This group also includes 'corporate reformers' that view AI as a means of instilling business-like efficiencies and market logics into education, aligning education systems with future workforce demands of 4IR and so on.
- **Social proponents of AIED**: people enthused by potential societal impacts of AI technology. This group tends to frame AIED in terms of ideas of 'AI for Social Good'- for example, making education more accessible, or meeting the diverse needs of learners.
- **Social critics of AIED**: people concerned by what they see as the societal misapplication of AI technologies – in particular where AI works to increase inequalities, disadvantage, discrimination and injustice. This group tends to have little/no involvement in the technicalities of AI development.
- **Educational critics of AIED**: educators concerned by the ways in which AI standardises and constrains teaching and learning. This group tends to judge AIED in terms of progressive education ideas, agendas around 'critical pedagogy', and other humanist understandings of education practice.
- **Technical critics of AIED**: An emerging group of computer science 'insiders' who are becoming critical of how their work is used across society. This group tends to frame AI in terms of fundamental ontological and epistemological tensions between the computational procedures and statistical modelling that drive AI, and the social realities they purport to represent.

Third, it is instructive to think through how these different interests and agendas have evolved over the past 40 years or so. For most of this time, it is probably fair to say that the development of educational AI has really only been of interest to a few technical and educational proponents – leading to AIED being a relatively settled and harmonious field of activity for most of its history. Indeed, the academic 'AIED community' still comprises largely of people best characterised as either technical or educational proponents, with a few crossing over into the corporate sphere as their products have been taken to market and commercially developed.

Looking from the outside, two shifts have taken place over the past ten years or so. First, was an influx of additional involvement in AI in education during the 2010s from social *and* corporate proponents. This diversified the nature of people advocating for AI technologies in education. While these 'late-comers' might have had little in common with technical and educational proponents, their increased presence in AIED seems not to have proven too disruptive given their underpinning desire to see more AI in education.

Latterly, however, the incursion of critical voices into this mix seems to have proven much more threatening to the harmonious nature of the AIED project. This has led to a number of different responses from AIED proponents. For example, some technical proponents have also tried to reframe criticisms as essentially technical problems that can be fixed. Other technical proponents have been quick to

dismiss critics as misinformed and lacking in basic understanding of the technologies that they address[1].

Elsewhere, some educational proponents have started to accuse critics of unfairly focusing on the application of AI in education systems and institutions, and therefore missing the wider potential for empowering individual learners who are self-regulating and sovereign in their choices and actions. Some social proponents of AIED are also beginning to accuse critics of excessive negativity – e.g. ignoring cases where AI technologies have proven to successfully support particular types of learners. At the same time, there is an increased willingness to blame the excessive hyping of AIED products by corporate actors who have latterly entered the area with different motivations. All told, for some insiders who have long been involved, AIED is starting to feel less consensual, less straightforward and, perhaps, a less appealing place to be.

Questions Raised by the AIED Backlash

So, how might the AIED community move on from this predicament? Here, I would like to suggest that it is important to resist any instinctive response to try to 'fix' these issues and propose 'solutions'. Neither it is productive to chide people for not focusing on positive cases, or else demonising the IT industry for taking AIED ideas, concepts and designs to market. In short, the points being raised by critics of AIED are not issues that can be simply dismissed out-of-hand, or resolved quickly.

Above all, it is important not to presume critics of AIED to be lacking in technical understanding, educational vision, and/or willingness to celebrate success. What if we do not presume that these criticisms are fundamentally wrong, but instead engage with what these criticisms are telling us about AIED, and why critics are increasingly feel the need to raise such issues? As such, I would like to argue that the points being raised by critics of AIED are worth unpacking in further detail. Here, then, are three starting-points for deconstructing what lies

---

[1] Editorial reviews to earlier drafts of this paper asked for references to the 'But, You Haven't Built AI' push-back provoked by critical commentaries of the technology. In short, this attitude can be seen on a daily basis to the responses on social media to prominent voices in the AI and ethics field - particularly if they happen to be women, non-white and/or not working for tech companies. From a personal perspective, here is a reviewer comment from a prominent AIED name to my 2019 'Should Robots Replace Teachers?' book:

"Those of us working in AI and Education and those in AI generally tend to form an early and strong understanding of how hard this stuff is to do. I looked up Prof. Selwyn's background and admire his career and contributions greatly. However, I noticed that (it seems) he hasn't attempted to build an AI system for education for himself. In my eyes (and those like me), this is a credibility problem. It is easier to have fears and to label negative outcomes as feasible when you haven't had experiences to draw from that may reveal the true challenges of achieving that scary goal. I was relieved in the latter parts of the book where the true challenges of teaching were listed and rightfully stated to be out of the scope of AI now and well into the future. But I'd still feel more comfortable if Prof. Selwyn took time to build some AI systems."

behind the AIED backlash, and therefore engaging with these arguments in a constructive rather than challenging manner:

i   What specific forms of AI in education are prompting push-back?

First are questions of specificity. Criticisms of AIED tend to be less broad-brush and more balanced than might first appear. Indeed, most critics of AIED are not *wholly* opposed to the presence of all AI in education, but raising specific criticisms of particular applications of AI. As such, it is useful to reflect on the types of AI in education that are provoking most push-back, and the particular problems that they are seen to be associated with.

For example, critics often call out forms of AIED being brought into schools and universities to monitor and track students. This includes the fast-growing uptake of online exam proctoring and plagiarism detection software. This also includes 'student safety management' systems that monitor student social media use, hall-pass apps that track student restroom visits, and other such 'spyware' that reframes surveillance as pedagogical care and safety.

Another common source of concern are AI applications that allow commercial interests to profit financially from public education – such as software designed to bring students into contact with targeted advertising, data-brokering, and other facets of surveillance capitalism. Alongside these egregious products are many other AI applications that might appear more benign, but critics might consider to detract from the ability of teachers to exert their professional expertise, diminish the social relations of the classroom, or otherwise contribute to the hollowing-out, dehumanisation of the educational processes and practices being automated through AI (e.g. Selwyn, 2022).

All these are examples of push-back against technologies that alter the conditions, mood and purpose of education. Many of these are instances of AI technology exacerbating existing characteristics of school and university systems that are based around standardisation, mass processing of students, monitoring, tracking and control. These are likely not to have been the intentions of those developing AIED technologies, but are certainly how these products are being experienced when implemented in real-life classroom contexts. As such, these examples highlight the need for proponents of AIED to think a little more deeply about the politics of how and where these technologies are being applied – i.e. the politics of the contemporary school or university, the politics of national education systems and education governance.

ii   What is motivating people to be critical of AI in education?

Second are questions of intent. Critics of AIED are usually not raising these issues simply to be obstructive, and neither are they claiming these to be universal truths. Rather, in pointing out such harms, critics of AI generally see themselves as acting as what Apple (2016) terms 'critical secretaries' – i.e. highlighting struggles that many minoritized people and groups are engaged in when AI is imposed into various aspects of their everyday lives.

In railing against particular forms of AI in education, critics are often attempting to point out social 'harms' (see Shelby et al., 2022). As is now widely recognised in various instances, some of the most substantial harms occur when AI models amplify the discriminations baked into their training data and subsequently drive AI systems to discriminatory and disadvantaging ends. Other harms might appear less substantial, yet also need to be taken seriously – for example, online exam proctoring systems failing to detect the faces of Black students, or non-binary students having to mis-identify themselves as either 'M' or 'F' in order to register a 'valid' system profile.

Such harms might not be designed deliberately into technologies, but arise from the inevitable reductions and losses inherent in representing complex social phenomenon in datafied forms, and then processing these data through complex computational procedures. All told, perhaps the most pressing matter to be discussing about the roll-out of AI in education is how this technology "has a tendency to punch down: that is, the collateral damage that comes from its statistical fragility ends up hurting the less privileged" (McQuillan, 2022, p.35).

Of course, these are all issues that usually remain completely out-of-sight for the vast majority of AIED developers and implementers. No-one sets out to develop technology with the explicit intention to 'harm' young people, and these are likely not to be issues that are personally experienced or directly witnessed by middle-class, well-educated and technologically-adept proponents of AIED.

Yet AIED critics are motivated by the belief that these harms associated with automation in education are relational in nature, and therefore likely to be experienced differently according to peoples' different backgrounds and circumstances. As such, what might be seen as a 'small harm' for one person will be far more substantial for others. In this sense, critics are reminding AIED researchers and developers that - regardless of their own positive experiences - they need to remain aware of the embedded socially-differentiated ways in which this technology can impact on others in harmful and deleterious ways.

iii　What values and ideologies are driving proponents' interest in AIED?

Third, are questions of personal values. Generally, then, the framing of AIED in terms of social harms is rooted in issues of power and politics – not something that often drives discussions in journals such as IJAIED. Indeed, many of the most critical concerns surrounding the (mis)use of AI in education are profoundly political in nature, and entangled with broader dynamics of power, disadvantage and marginalisation (see Verdegem, 2021). More specifically, is the concern that the forms of AI beginning to pervade education tend to "skew strongly toward the centralisation of power" (Crawford, 2021, p. 223).

In this sense, critics remind us that AIED needs to be approached as a political project. Reframed in this light, there are a number of ways that future discussions around AI in education might progress. For example, everyone working in AIED needs to be ready to examine (and to make explicit) the underpinning values and ideologies that are driving debates around particular issues. This involves reflecting on one's own positionality, as well as pushing back against any claims for AI to be non-political and neutral. Indeed, as Green (2021) reasons, the tendency for technologists to claim neutrality for their work is a 'fundamentally conservative' position

that constitutes tacit support for maintaining the status quo and, therefore, the interests of dominant social groups and hegemonic political values.

This raises various points of reflection that the AIED community needs to address head-on. For example, how complicit are you prepared to be with the commercial marketing of AIED products, and/or the ways that AIED is being deployed in school and university systems? While AIED proponents might like to imagine their work solely in terms of supporting individuals to make rational learning choices in 'personalised' and 'agentic' ways, this does not carry over when technologies are being deployed in education institutions and marketplaces that are built around decidedly different values.

Conclusions

Underpinning all these points is a call for proponents of AIED *not* to engage with criticisms of their work in an overly defensive and/or combative manner. One of the obvious failings in contemporary debates around AI is the tendency for AI specialists to move quickly onto the front foot and attempt to shut down criticisms of their work – belittling critics and asserting their expert status (e.g. the perennial taunt that "you cannot critique AI until you have built AI"). Some critics are now beginning to quite reasonably answer back that AI proponents show little understanding of social issues. Entrenched positions of 'them and us' are then taken up, and mutual hostility ensues.

Another obvious failing in general debates around AI is the tendency for AI specialists to want to reframe the concerns of critics on their own terms – engineering forms of 'explainable AI', fully-representative training data, ethics codes, privacy pledges, or commitments to 'fairness, accountability and transparency'. While well-intentioned, none of these changes do much to address the fundamental questions of politics, power, institutionalised injustice and social harm that lie at the heart of the current AI backlash.

Yet, for the time being, the AIED community has an opportunity to *not* fall into these same traps. From this point onward, then, it would be good to see AIED proponents *and* AIED critics start working together to explore differences of opinion and find points of agreement. Returning to the initial rhetoric arising from the panel discussion at the AIED 2022 conference in Durham, there is a need for the academic AIED community to carry on its work … but not carry on 'regardless'. All told, if we *are* currently in the midst of an AIED 'Coming of Age' then hopefully this will take the form of a sophisticated political awareness being co-produced by proponents *and* critics of this technology.

References

Apple, M. (2016). Challenging the epistemological fog. *European Educational Research Journal, 15*(5), 505–515.

Atkinson, R., Brake, D., Castro, D., Cunliff, C., Kennedy, J., McLaughlin, M., & New, J. (2019). A policymaker's guide to the "techlash". Information Technology and Innovation Foundation. October 28th, https://itif.org/publications/2019/10/28/policymakers-guide-techlash/

Birhane, A., Kalluri, P., Card, P., Agnew, W., Dotan, R., & Bao, M. (2021). *The values encoded in machine learning research*. https://arxiv.org/abs/2106.15590. Accessed 15 June 2023

Bowles, N. (2019). *Silicon Valley came to Kansas Schools. That started a Rebellion* (21st.). New York Times. www.nytimes.com/2019/04/21/technology/silicon-valley-kansas-schools.html. Accessed 15 June 2023

Crawford, K. (2021). *Atlas of AI*. New Haven, CN: Yale University Press.

Edelman, S. (2018). Brooklyn students hold walkout in protest of Facebook-designed online program. New York Post, November 10th, https://nypost.com/2018/11/10/brooklyn-students-hold-walkout-in-protest-of-facebook-designed-online-program/

Green, B. (2021). Data science as political action. *Journal of Social Computing,2*(3), 249–265.

Harwell, D. (2022). Cheating-detection companies made millions during the pandemic. Now students are fighting back. In Martin, K. (ed.). *Ethics of data and analytics* (pp. 410–417). New York, Auerbach

McQuillan, D. (2022). *Resisting AI*. Bristol: Bristol University Press.

Schleicher, A. (2021). *How smart can education get? Very smart*. June10th, https://oecd.ai/en/wonk/digital-education-outlook-2021

Selwyn, N. (2022). Less work for teacher? The ironies of automated decision-making in schools. In S. Pink, M. Berg, D. Lupton, D. & M. Ruckenstein (Eds.), Everyday automation: Experiencing and anticipating automated decision-making (pp.73–86). London, Routledge

Shelby, R., Rismani, S., Henne, K., Moon, A., Rostamzadeh, N., Nicholas, P., Yilla,N., Gallegos, J., Smart, A., Garcia, E. & Virk, G. (2022). *Sociotechnical harms: Scoping a taxonomy for harm reduction*. https://arxiv.org/abs/2210.05791v1. Accessed 15 June 2023

Vallor, S. (2022). We used to get excited about technology – what happened? MIT Technology Review, October 21st, www.technologyreview.com/2022/10/21/1061260/innovation-technology-what-happened/

Verdegem, P. (2021). *AI for everyone*? University of Westminster Press.