23. Imaginaries of responsible Generative AI in secondary education

Jen Ross, Esther Priyadharshini, Harry Dyer, Ayça Atabey, Colton Botta, Cara Wilson and Judy Robertson

INTRODUCTION

Imaginaries of educational Artificial Intelligence (AI) are influential in policy and practice, but the voices and perspectives informing them are limited and limiting for making productive interventions that reflect the values, hopes, and concerns of young people. This chapter presents findings from a research project funded in 2024 to scope the potential for 'bridging Responsible AI divides' in the formal education sector in the UK. These divides are understood as arising because discussions of Responsible AI (RAI) tend to be conceptual and high-level, and to take place outside of specific contexts of policy and practice where RAI will have to be legislated, implemented and evaluated. The projects funded in the initial scoping stage of a larger programme were tasked with defining 'what Responsible AI looks like across different sectors of society'. Our project examined RAI in secondary education, and how the perspectives of young people in schools might shape discussion and decision-making about AI futures. We used arts-based approaches to engage young people about their hopes, concerns, and imaginations for the future, with a focus on Generative AI (GenAI). GenAI was chosen because of its unique position in the education landscape at the time. First, its development and rapid adoption was being driven from outside the sector, and it was being used in ad hoc and 'bottom-up' ways by students, teachers, parents, and the public. Second, GenAI's capabilities, and predictions about its future ones, were raising questions about what should be valued in the education of humans in an age of machine intelligence, and it was therefore implicated in how educational cultures and subjects were being conceived.

Our work took place against a backdrop of intense debate about the nature of education, and especially assessment, at a time of emerging impacts of the rapid adoption of GenAI. This came in the wake of the release of high-profile commercial platforms and tools for generating novel text, images, and other material in response to user-generated prompts. GenAI was controversial from the outset: because of the types of tasks people were using it for, its use of source material without permission, its tendency to replicate dominant perspectives and biases, and the huge energy demands it made. In educational settings, people were debating these issues as well as GenAI's potential role in augmenting or even replacing key educational processes, such as demonstrating knowledge and understanding through writing, creating educational materials and assessments, answering questions and explaining concepts. For these reasons and more, GenAI was the right focus for an exploration of RAI in education. To focus the project's work further, we concentrated on three principles of RAI of special relevance to education: explainability, fairness, and privacy (Blackman, 2022).

In this chapter, we analyse and discuss what we learned from a series of workshops with a total of 22 young people in both mainstream and additional needs education settings. The workshops took place over a month and investigated young people's understanding of AI and GenAI, their approaches to creatively engaging with GenAI tools, and the different ways they imagined responsible and irresponsible AI in the future of education. We identify and discuss two key findings. First, despite bringing well informed, imaginative, and critical perspectives to bear on a range of matters relating to GenAI, young people generally did not apply these perspectives to critique or reimagine schools or schooling in a current or future age of pervasive AI in Education (AIED). Second, their orientation to AIED futures was one of conserving and preserving education, protecting what they saw as its foundational qualities – hard work, trustworthiness, and detachment from the extractive, datafied world. We explore the significance of these findings for navigating risks in RAI work in educational settings. First, there is a risk of responsibilising young people inappropriately by suggesting that dealing with the significant pedagogical, ethical, and safety implications of GenAI should be a matter of their choices and literacy. Second, there is a risk of negatively affecting young people's trust towards education and schooling as a place that will protect them, their rights, and their individuality from negative consequences of GenAI that they see in wider society.

GENERATIVE AI IN EDUCATION

Responsible AI

The growth of AI has led to calls for better oversight and responsible leadership to minimise the harms that these technologies can create. Alongside policy-driven attempts to define what 'Responsible AI' might look like (European Commission, 2019) a number of Big Tech companies making investments in developing AI have set out their own principles for what they consider to be 'Responsible AI'. For example, Microsoft's (2022) requirements for AI suggest six guiding principles of AI: fairness, reliability and safety, privacy and security, inclusiveness, transparency, and accountability. However, what these principles mean in action, who they matter to and for, how they are experienced in everyday interactions with AI, and where the responsibility for upholding these principles falls are points of significant debate both in education and in wider socio-political contexts.

Companies have adopted rhetoric around principles such as explainability, fairness, and privacy, but a lack of public trust in this rhetoric shifts the responsibility of upholding these principles to the general public. As the Pew Research Center (McClain et al., 2023) notes, 70 per cent of US adults surveyed said they have little or no trust in companies to make responsible decisions about AI and 81 per cent said that the information collected through AI will likely be used in ways they are uncomfortable with. While 78 per cent trusted themselves to make the right decisions about their data, 61 per cent of them were sceptical that any action they take will make much difference – an example of what Draper and Turow (2019) have described as 'digital resignation', or what Hargittai and Marwick (2016), in the context of young Internet users, call 'online apathy'.

AI designers also struggle when designing systems with concepts such as fairness in mind, particularly as AI systems tend to reflect extant inequities and complexities baked into different systems. For example, Obermeyer et al. (2019) note of health-related AI algorithms:

> Racial bias reduces the number of Black patients identified for extra care by more than half. Bias occurs because the algorithm uses health costs as a proxy for health needs. Less money is spent on Black patients who have the same level of need, and the algorithm thus falsely concludes that Black patients are healthier than equally sick White patients. (p. 447)

In this manner, algorithms designed to ease burdens in different systems end up further harming marginalised communities. Fixes for biases in AI also pose concerns and problems. For example, research suggests that attempts to control toxicity in language models (LMs) 'comes at the cost of reduced LM coverage for both texts about, and dialects of, marginalised groups' (Welbl et al., 2021, p. 1). Welbl et al. (2021) go on to note that designers aim to 'solve' the racism of AI models through removing discussions of marginalised groups altogether. For this reason, discussions of 'Responsible AI' need to consider how design choices are experienced by different stakeholders, and how users are increasingly placed in processes and positions of responsibility.

Educational implications of AI systems are similarly fraught. Principles of explainability would suggest that an AI system deployed in a school setting should be able to explain the decisions it makes and the information it produces with truth and clarity, tailoring the intelligibility of the explanation to the stage of the learner. However, current GenAI models do not generate explanations by default, so the content they produce cannot be fully scrutinised, nor the models adapted, and the information produced by GenAI is not guaranteed to be accurate, so the task of critically engaging with its outputs falls on teachers and students (Miao & Holmes, 2023). AI decisions in education should be fair and avoid bias, but defining the priorities for balancing different definitions of fairness is complex (Kearns & Roth, 2019), and in any case, current approaches to training large language models may result in outputs that replicate bias (Salinas et al., 2024; Whittaker et al., 2019). Issues of consent, data rights, and ethical data collection have not been resolved in relation to GenAI. As Dignum (2021) observes, dealing with conflicts of interest around educational AI will be a significant policy and legislative task. At present, there is little indication that this task is well understood or being taken up in educational systems.

GenAI in Education

With the introduction of ChatGPT in late 2022, the rapid emergence of, and increased public access to, GenAI has sparked significant interest and debate in the field of education. Critiques of GenAI include that it is enabling widespread cheating as students use it to produce texts and other material that they then claim as their own, changing the nature of teaching and placing new skill development demands on teachers (Jensen et al., 2024). Other analysis identifies its role in spreading misinformation, systemic bias, and prejudice (Selwyn, 2022), and damaging trust between teacher and student (Luo, 2024). More general fears about the loss of human creativity and originality are also seen as significant in education (Higgs & Stornaiuolo, 2024).

Nevertheless, there is evidence of considerable uptake of GenAI tools by both learners and teachers. For learners, uses include brainstorming ideas for essay writing and creative projects, breaking down complex tasks into simple steps, and exploring academic interests that their teachers may not cover (Higgs & Stornaiuolo, 2024). For teachers, these include translating learning material (Yeh, 2024), developing lesson ideas and providing feedback on the quality of lesson plans before students see them (McDonald et al., 2024). One GenAI tool,

ChatGPT, has been shown to generate both multiple-choice and free-response questions that are relevant and match the specified difficulty and learning objectives for assessments (Onal & Kulavuz-Onal, 2024).

In parallel, significant work and investment are being directed towards formal educational applications of GenAI. These applications are largely promised for the near future (and echo promises that have been made for decades), and include:

- personalised tutoring, for example to clarify questions and give assistance outside of normal classroom hours (Higgs & Stornaiuolo, 2024; Su & Yang, 2023).
- time-saving for teachers in the form of producing learning content, lesson plans, assessments, and feedback (McDonald et al., 2024). Work is ongoing to develop GenAI applications for marking free-response questions, providing scores and written feedback.

Big Tech players are developing and disseminating their preferred educational perspectives and approaches. For example, Google's research team envisions a future world in which GenAI offers 'a personal tutor for every learner and a teaching assistant for every teacher' (Jurenka et al., 2024, p. 1), though they note this is far from being the current reality in classrooms. A prevailing theme is that GenAI should not replace teachers. Instead, teachers should partner with AI (Van Den Berg, 2024).

These kinds of claims require a substantial knowledge base in order to engage with them critically, and the state of AI literacy in schools is therefore extremely relevant to the situation of GenAI in and around education. In the absence of strong and overarching policies and practice guidelines in most education systems, understanding the current state of knowledge around AI amongst students and teachers has become crucial (Antonenko & Abramowitz, 2023; Sperling et al., 2024), along with addressing misconceptions and myths about AI (Bewersdorff et al., 2023). Such myths have been studied in recent work exploring children's beliefs about AI. Kim et al. (2023) find that middle school students often hold naive preconceptions of AI, such as equating it with automation or believing it to be impartial and fair. Similarly, Andries and Robertson (2023) find that primary school children in Scotland lack an accurate understanding of data privacy and security in relation to conversational agents.

AI literacies are addressed through different lenses, including technical, ethical, societal, and rights implications of AIED. They also involve understanding policy and legal frameworks that place responsibility on companies providing GenAI tools, and on schools or local authorities who deploy such tools, to ensure transparency and inform stakeholders, particularly when groups like younger children are involved (Linderoth et al., 2024). Dai et al. (2024) present an 'embodied, analogical, and disruptive' approach for upper elementary students, focusing on human-AI comparison through analogical teaching and embodied interaction. Additionally, Domínguez and Stoyanovich (2023) advocate for 'responsible AI literacy', emphasising a stakeholder-first approach that addresses diverse audiences and considers the ethical and social implications of AI. Lao and You (2024) stress the importance of GenAI literacy to involve understanding, using, and evaluating GenAI while addressing its potential harms. Regardless of the specific focus, the onus is usually on the knowledge and understanding of individuals, including those with minimal power or influence to effect systemic change. This might usefully be described as 'responsibilising' RAI in education.

Responsibilisation and Education

Formal education in the UK and elsewhere has been shaped by forms of thinking and governance that position teachers and students as *individually responsible* for their futures. This tendency has been described as 'responsibilisation': 'a governance praxis that operates through ascribing freedom and autonomy to individuals and agents … while simultaneously appealing to individual responsibility-taking, independent self-steering and "self-care"' (Pyysiäinen et al., 2017). Coined as a term in the late 2000s (Shamir, 2008), responsibilisation has been broadly studied and critiqued in educational settings. Researchers have examined issues including assessment (Torrance, 2017), datafication of student behaviour (Whitman, 2020), parental choices in schooling (Liu & Bray, 2022), and lifelong learning (Regmi, 2023) through this lens. In their work on university ranking initiatives, Decuypere and Landri (2021) trace the impacts of standardisation and personalisation of university choice in terms of responsibilisation: 'what is of value for oneself (and what is not) and what matters for oneself (and what does not)' (p. 11) – this, they argue, embeds the idea that 'making right choices becomes the sole responsibility of users themselves' (Decuypere & Landri, 2021, p. 12).

Responsibilisation has not yet been explored in the context of AIED, though a legal paper from Torres (2023) introduces the concept of 'moral responsibilisation' in RAI (p. 150), referring to the need to go beyond legal demands to ethical dimensions of producing AI systems. The author helpfully connects this to practices of innovation and care, but does not analyse the problematic nature of responsibilisation itself. Responsibilisation as a critical concept has relevance to thinking about RAI, particularly in light of critiques of the lack of clarity about *who* or *what* is 'responsible for' RAI (Coeckelbergh, 2020; McGrath et al., 2023). Responsibilisation offers a way in to critically exploring RAI, because it highlights the risks of failing to designate responsibility in meaningful, specific, and appropriate ways: that individuals in the education system will come to see AI as another set of risks (among many) they must navigate, as part of demonstrating their 'willingness to comply with the process and compete for the rewards' (Torrance, 2017, p. 91) of education.

RESEARCH DESIGN AND METHODS

This chapter draws on insights from a work package within a larger study. The aim was to explore young people's experiences and perceptions of GenAI. Through creative workshops, we aimed to investigate how developments in AI impact how young people value education and thus how educational cultures and subjects may be emerging. To do this, we explored how young people envisioned futures for AI in their lives and in learning and educational contexts. Using GenAI applications in our workshops, we also tried to understand the nature of their engagement and creative outputs with these tools. We did this using speculative methods to imagine different futures. Ross (2022) explains that 'a speculative approach works with the future as a space of uncertainty and uses that uncertainty creatively in the present' (p. 13). Speculative futures work has increasingly gained attention across disciplines like learning sciences and human-computer interaction, particularly in envisioning AI's role in education. Dunne and Raby (2013) address the role of design as a speculative tool, asking 'what if' questions to explore desired futures, where young people's imaginations can inform technology design. Speculative design work often involves participatory approaches, as efforts are made

to empower young people to shape future technologies (Bai et al., 2023; Hossain & Ahmed, 2021; Kenny & Antle, 2024).

The work package involved young people at three sites – a secondary school in Edinburgh (young people aged 16–17), another in Norfolk (ages 17–18), and an additional needs school in Edinburgh (ages 13–16). In all, 22 young people participated in four workshops per school, each workshop lasting about 90 minutes. As abilities varied in the additional needs school context, two separate sets of workshops were conducted there. In total, 16 workshops were completed. At each site, researchers collaborated with artist-practitioners to design a series of workshops that incorporated imaginative, exploratory, and speculative activities. The facilitators of the additional support needs (ASN) workshops were chosen for their experience in conducting creative and participatory work with young people with ASN. Including and centring neurodiverse and ASN participants from the inception of this work was part of our commitment to RAI practices, to which fairness, accessibility, and inclusivity of AI systems are paramount. In this chapter, we have focused on perspectives from the young people in our study who were verbal – the significance of the work of non-verbal young people in our study will be discussed in future publications.

We were aware of the intense debates on the use of AI for assessed work and the questions this raised about the nature of learning and assessment. We also realised that working with young people within the space of their schools could encourage responses that aligned with the dominant discourses and regimes of schooling. Thus, it was vital that our methods in this project encouraged thinking-making activities, and discussions that were able to promote the speculative leap of imagination away from the everyday. We used arts-based approaches – such as sketching, mapping, collaging, creative writing, and crafting – as a way of encouraging speculation, imagination, and playfulness (Auger, 2013; Lukens & Disalvo, 2011; Malinverni et al., 2014). Some young people created maps of their everyday technology use (including AI); some tried to create sketches/collages of what they thought GenAI looked like; some experimented with GenAI tools/applications by entering prompts to chat/converse with the application (ChatGPT or Copilot) or to create visual outputs (Fulljourney) and then discussed their experience; some played with MindsEye to find ways of representing themselves. Other activities involved creative expression through collage; and placards of protest about an imagined future state of AI. These activities were offered as ways of sharing ideas about GenAI that may be hard to imagine or express through more traditional, language-reliant methods (Lyon & Carabelli, 2016). Another intention was to encourage the safe use of emerging technology in creative ways to allow young people to pick up useful skills in manipulating software while engendering a critical or questioning relationship with such technology. For each site, we adapted the workshop structure to remain responsive to the emerging needs or curiosities of young people.

The research plans were submitted and approved by the research ethics committees of the University of Edinburgh and University of East Anglia, and paid particular attention to the security of young people's information and the storage and use of their creative digital outputs produced using specifically subscribed GenAI tools. We discussed the importance of not offering participants' personal information to AI tools; we encouraged the use of textual (not visual) prompts; all outputs were securely stored; and subscription services to AI tools allowed us to retain better oversight of inputs/outputs generated by young people. Across the workshops, the use of these tools prompted discussions of privacy, fairness, accountability, accuracy, security, creativity, entertainment, and learning. The workshop series culminated

with a reflective session on participants' main takeaways on GenAI, through the co-creation of 'zine' materials (Hay, 2022). This creative activity was intended to support the sharing of insights from the work in young people's own creations rather than solely through researcher-led summaries. The final zine (Atabey et al., 2024) was disseminated in physical and digital form to participants and key stakeholders (schools, policymakers, academics, and industry).

Analysis

The analytic process for engaging with these materials was iterative, involving the images and participant-produced materials, observation field notes taken by the researchers during each session, and parts of audio recordings made during the workshops. We shared and refined observations through individual note-taking, discussions, collaborative annotation of images, and debates about a range of insights and curiosities that emerged from each site. The interdisciplinary nature of the team (representing sociology, design, computer science, law, and education researchers) created opportunities for refining our thinking through confronting theoretical and methodological differences within the group.

Our analytic conversations often focused on salient issues, surprises, and what MacLure (2013) describes as data 'glows'. These were aspects in the workshops or materials that drew us closer because they were intriguing for some reason. Sometimes these data glows centred on widespread perspectives shared by a large number of the participants across sites (e.g., around education as 'effortful', as discussed below). At other times, we found salience in expressions of affect and emotion, or the urgency and intensity of discussions (either in the workshops or amongst the researchers). This approach to analysis supports the nature of the experimental and playful workshops that generated unusual data including creative outputs, curious questions, playful speculation, and an excess of affect and meaning that was impossible to routinely list, code, reduce, and manage (MacLure, 2023; St. Pierre, 2021).

Our focus in this chapter is on the tensions emerging from participants' critical engagement with GenAI in general, and their more trusting engagement with the concept of school/education and GenAI's role within it now and in the future. We move on now to this discussion.

FINDINGS

Young People's Perspectives on GenAI

Young people in our workshops were already familiar with GenAI and had reasonably accurate understandings of what it could do and the kinds of uses they and their peers were making of it. Their experiences largely echoed findings from the recent YoungScot survey of 2014 young people (YoungScot, 2024): to ask questions, for fun, to save time, to find out the right answers to things, to summarise information, to do school/college work for them (19 per cent of survey respondents selected this option), and to chat (p. 22). Nevertheless, the perspectives of the young people in our study should be understood as indicative rather than fully representative of the kinds of experiences, hopes, and concerns that young people have in relation to GenAI.

Aspects of the workshops involved exploring AI myths, understandings, and uncertainties, which helped establish a baseline for discussions and activities. For example, at the start of

one workshop, participants were asked to list some ideas about AI that they were aware of, and discuss their accuracy – ideas included that AI will replace every job or take over the world, and that there are legal limitations on what it is allowed to say. However, using arts-based methods in our creative workshops helped us move beyond practical and factual discussions of GenAI to elicit more metaphorical and imaginative understandings of AI and its effects. This allowed the participants to collectively build critical representations that reflected their concerns about AI and the 'costs' of its presence inside and outside of the classroom. These concerns centred on the *unknowability of AI*; its *untrustworthiness*; and its *social and environmental costs*, including enabling bad human actors to amplify their influence.

Unknowability of AI

Far from a strict definable force in their lives, the young people discussed the unknown repercussions of GenAI, existing as an ominous mass whose size, limits, and form remained mysterious and hard to grasp, looming in the distance (see Figure 23.1) – AI as a 'blob' or a spherical mass.

Despite their confidence in defining and describing GenAI, sometimes at quite sophisticated technical levels, AI was seen as unpredictable in both form and effect, both beautiful and dangerous, with unknown effects. For instance, when discussing AI as a beautiful sphere, the young people noted that this was in part how AI 'sold' itself, but that when you dug beneath this initial outside layer, there was an ominous and more dangerous blob underneath.

This 'unknown' nature of AI also came through in uncertainty about how well the outputs of AI could be trusted. For example, in Edinburgh, the young people discussed how GenAI may generate 'solutions' without offering suitable and transparent explanations as to how those solutions were achieved, characterising AI as responsible if it 'fails to properly help the user grasp concepts they wish to learn; gives solution without explanation; gives unreliable responses' (from participant writing during a workshop). They discussed how they may be asked to trust the work of the AI without seeing the steps it took to get to the results. Young people wanted to 'see the working out' of the AI to better trust the outputs it produced – endorsing the importance of implementing the RAI principle of explainability in GenAI. In a related set of discussions, there were concerns around the 'ease' of AI which removed the 'struggle' of thinking for themselves. See the discussion section later for more on this.

Untrustworthiness of AI

Worries about the reliability of GenAI also came with concerns about how unwaveringly certain GenAI was in its responses, and what room this left for distrust and criticality. One image in Norwich depicted how AI was forcing users to follow its outputs unquestioningly. This could lead to being asked to follow AI into unknown futures (see Figure 23.2).

In Norwich, the young people expressed their concerns around the ethics of AI, noting that they as consumers did not necessarily know the intent of the companies that created the technologies. They discussed the censorship that they felt AI enforced, and that they were not sure they could trust the designers when AI tools avoided answering questions around potentially controversial or sensitive issues. A particular topic of discussion was that AI presented itself as 'neutral' and free of bias, when the social biases of the humans that designed the AI were readily apparent. When playing with GenAI, the young people were often frustrated that the AI would not answer certain sensitive or challenging philosophical questions (e.g., 'the trolley

Source: Authors' own.

Figure 23.1 AI as spherical blobs, beautiful and dangerous

problem'[1] in Edinburgh), but that nonetheless they felt there were clear biases in the answers and outputs produced.

Other discussion emerged around the ways in which AI presented itself as seemingly human, but that what they observed was a facade of humanity. GenAI was often discussed as attempting to pretend to be human (intentionally or not), but its representation of humanity was both creepy and, at times, laughable.

Social and environmental costs

Many participants thought that often the 'real threat' was not GenAI itself, but humans – both those who designed it and those who would use it maliciously. For example, in Edinburgh, some of the young people explored the idea that irresponsible GenAI companies could sell

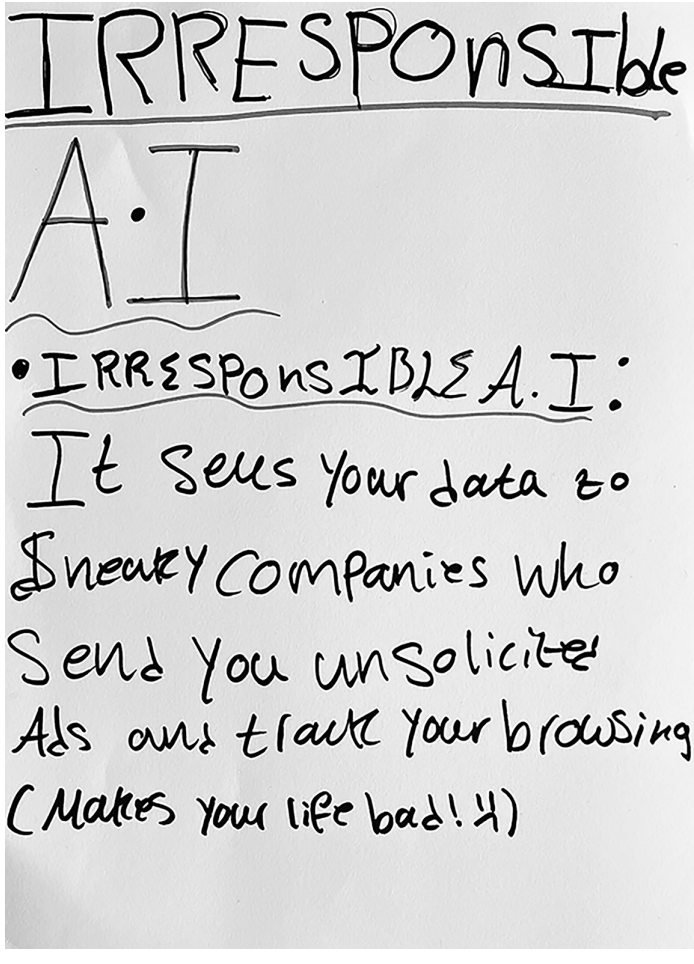
Source: Authors' own.

Figure 23.2 Do as we say!

their data to 'sneaky' companies, who would use it for surveillance and advertising, in turn making the lives of young people 'bad' (see Figure 23.3).

Worries and concerns about other social costs of AI included discussions around the ethics of AI, and the environmental costs of using it. Various pop culture figures and moments were highlighted, with the young people clearly aware of AI's presence in the cultural landscape around them. For example, discussions were raised about AI 'deepfakes' of Taylor Swift, a viral AI-produced public event in Glasgow titled 'Willy's Chocolate Experience', or the potential impact of AI on the UK general elections – all live social and cultural issues at the time of the workshops (mid-2024). This led to the young people playfully remixing political figures in their collages, reflecting the perceived silliness of AI through images such as an AI-generated King Charles's face on a cat (see Figure 23.4).

The young people also articulated worries and concerns about the environmental costs of GenAI technologies, for example, how much water/energy was used in the production of AI images. Discussions touched on the trade-offs they were asked to make when using AI, and what was lost and gained through these technologies. For example, opting out of using these technologies could mean missing out on personal benefits, but these benefits had to be weighed against the environmental and social costs of using AI. AI was also seen as somewhat inevitable, with young people having limited means to resist or avoid it in their lives.



Source: Authors' own.

Figure 23.3 AI selling data for advertising and surveillance

Young People's Perspectives on Education

We structured and designed the workshops, in part, to try to avoid making assumptions about how young people would see the relationship between AI and education. Dominant concerns in the educational literature, as discussed above, included a need for a radical overhaul of teaching and learning, the potential for widespread cheating, and loss of human creativity. Young people's perspectives on GenAI and education included discussion of all of these concerns, but not necessarily in ways the literature would suggest. For instance, in participants' envisioning of educational futures, many of the structures and relationships that make up education were not imagined differently, and there was no discernible discourse of educational

Source: Authors' own.

Figure 23.4 The cat king

transformation through AI. Many features of schooling were assumed to be enduring. We found that participants in our workshops did not see AI transforming the structures and relationships of contemporary education, as the examples below illustrate.[2]

The value of hard work

Young people were largely positive about ways AI could be a positive support in their learning, correcting mistakes and improving their writing, and allowing them to access relevant information in a convenient way. Efficiency was prized, with one participant in Edinburgh noting: '[ChatGPT] *c*an pull resources from, like, everywhere … not take hours and hours and

hours'. This was balanced by the kinds of concerns highlighted in the previous section about accuracy and bias: 'The bad thing is you don't know where I'm getting that stuff from. You don't know what data it's been given in the first place.' Another participant asked: 'Would you feel confident presenting that or would you want to check?'

However, young people thought that improper or excessive use of AI could become a detriment to learning, which ought to involve a degree of effort from the learner. There was widespread scepticism that using AI to do coursework (as opposed to supporting people to do it themselves) would be helpful in the long run. Discussions about AI providing potentially 'dubious' shortcuts on homework tasks or coursework emphasised beliefs around the value of hard work and moral or responsible learning behaviour. Groups discussed how taking shortcuts may encourage lazy habits that would hinder learning and let them down in an exam situation. There were also discussions about the risks of getting caught, as well as the sharing of anecdotes about people who managed not to get caught using GenAI to produce schoolwork.

Futures for GenAI were assumed to include the possibility that it could keep improving to the point that it becomes impossible to tell its outputs apart from solely human effort. In this situation, young people were less inclined to imagine an overhaul of education systems than a redoubling of efforts to preserve traditional assessment by conducting it all under supervision or in exam conditions. This may reflect an implicit trust in teachers and schools to assess stringently and in a manner that supported young people's future development; and also a sense that examination is so central to education as it is currently practised that this would never be allowed to change.

Safe EdTech in an unsafe tech world

Most young people were highly critical and cynical about dubious practices by Big Tech companies, and many conversations revealed low levels of trust towards them and their uses of and priorities around GenAI. However, when it came to exploring GenAI as an integral part of schooling in the future, there seemed to be an unspoken assumption that any technology used within an educational setting would have been rendered safe and acceptable. While there was scepticism about age-based filters to protect younger users, there was a consensus that the use of AI within schools would and should be safe: 'So if AI is made safe in a way that it can be used with schoolchildren or in a school environment, then it's fine. So like people who use it maliciously, that's different, that's sort of different AI' (workshop participant, Norwich).

Such sentiments prompt us to consider how education and school conjure – at least for some young people – a sense of a 'safe space' and associated feelings of security and protection. The high degree of ethical standards around digital content and practices that young people associate with education may reflect their familiarity with stringent safeguarding and privacy efforts in schools, including protecting children from seeing harmful content (as exemplified through Internet filters within school grounds). Perhaps discourse of school as a well-protected digital space engenders an assumption that AI technology that enters this space will also be stringently monitored to ensure young people's safety. We noticed that even the most critical discussions of a datafied world 'out there' were not applied to schools. There were no discussions of the datafication and surveillance practices embedded within contemporary schooling itself. In short, critiques of AI were not critiques of education or formal schooling. This prompts us to ask: What kind of work is done by the adjective 'educational' in educational technology? Does the prefix 'educational' foster a sense of security

that acts as a barrier to critical questions of the kind young people are asking about GenAI in the wider digital world?

DISCUSSION: TROLLS UNDER THE BRIDGE – THE RISK OF BRIDGING THIS RESPONSIBLE AI DIVIDE

The programme of funding that supported our project framed a 'divide' between RAI concepts and the implementation of RAI in practice. Our project aimed to build a bridge that would enable young people's perspectives to become part of decision-making about AIED. In doing so, we discovered a metaphorical troll under this particular bridge – a risk of sparking cynicism about schools and education by encouraging critical perspectives about a setting young people currently appear to feel some trust and optimism towards. Selwyn (2022) warns that speculative approaches to AI, whether determinist or critical, can sometimes distract from engaging with the 'actual substance' of AI technologies, particularly when speculation is driven by industry claims that are not fully realised. However, we found young people ready and able to engage speculatively and critically about this actual substance – but not to apply this to their thinking about education futures.

Cynicism, or what has been described as 'digital resignation' (Draper & Turow, 2019), comes about when risks and problems with technology are understood, but where agency to act and change things is limited or not perceived to be effective. For all the talk about the importance of young people's voice and rights, in educational settings young people have little meaningful ability to withdraw consent for the use of technology. We began to consider: What is the value for young people themselves of fostering critical AIED literacy in schools? Does it risk responsibilising them, if activated in conditions where this knowledge cannot lead to change?

This question was confronting for us, as a number of us work with participatory design and futures methods and approaches, children's rights framings, and other ways of valuing and including the voices and imaginations of young people in shaping the future of education. Despite these commitments, we are not sure that it is responsible at this time to ask young people to engage with the range of tensions and issues that AIED is producing. These issues include the levels of uncertainty education leaders are experiencing around AI, the unanticipated impacts of platformisation and datafication on teaching and learning, the pressures on education systems to authorise for-profit AI and other tech interventions in schools, and the whole edifice of practice, policy, and relationships that, as critical AI and education scholars, we seek to problematise.

Furthermore, young people are not, by and large, coming to questions and critiques of AIED on their own. The young people in our study are savvy AI consumers in their personal lives, and insightful critics of AI in the contexts of creativity, sustainability, and other domains. They experience demands on them to navigate very risky AI practices outside of school – for example, to be experts in sifting fact from fiction, and tread the line between 'tool' and 'replacement' for their efforts. They are already facing the limits of their capacity to change and influence the direction of AI, with humour, realism, and some flashes of anger. Perhaps they do not have the appetite or energy left for scepticism when it comes to education. Or perhaps they are resolving their lack of agency by choosing to trust. We heard from a

number of young people about the need to preserve school as a 'safe' space where AI is always used responsibly.

In other words, while young people do not believe that Big Tech will protect them, their rights or their individuality, they *do* believe that schools can and will erect boundaries on their behalf, to preserve their classrooms as a place of respite from the need to be so vigilant and critical elsewhere. This challenges the principle of critical AI literacy for young people as a pathway towards more responsible AI. Young people's imaginations of education futures are a crucial resource to inform ethical advocacy about the implementation and governance of GenAI. However, if there is no clear route towards changing or improving what is unfair, exploitative, or opaque in the implementations of AI in schools, then the education sector may not be ready for young people to critically confront the tensions and irresponsibilities of AIED or to offer resolutions. This suggests the need for an approach to AI literacy that treads a delicate line between awareness raising and attending to relations of power where schools, policymakers, researchers, and education leaders need to create better conditions for young people's voice and activism towards Responsible AIED. That line is already being explored in related contexts – for example, in the implementation of the United Nations Convention on the Rights of the Child (UNCRC) in Scotland, where it became law in January 2024. Schools are required to teach the law accurately and respond appropriately if young people feel their rights are not upheld at school, with organisations such as Enquire[3] able to advocate for children and families where this is not the case.

WHAT NEXT?

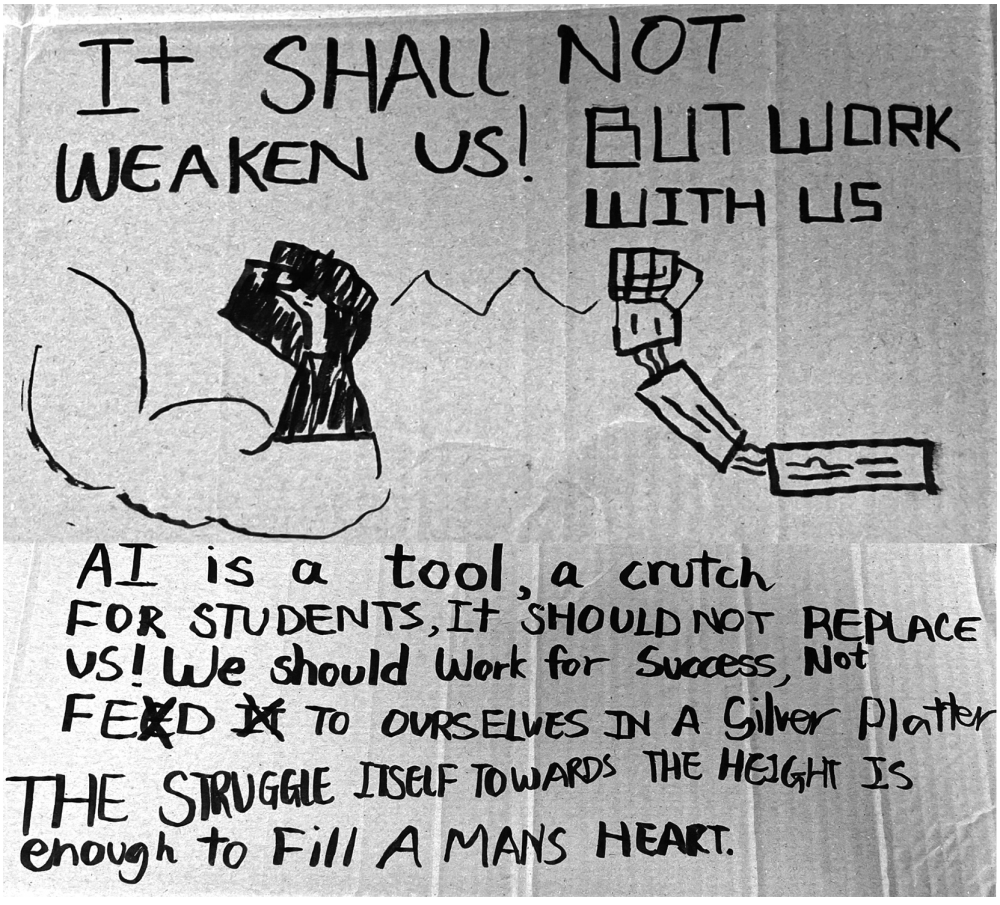
In one of the speculative workshops, young people were asked to imagine what schools, learning, and AI technology would look like and do in a far future (2074). They engaged in a creative 'demands for Big Tech' exercise, which resulted in a set of banners and posters articulating imagined future students' concerns and desires in relation to AI.

Of the demands that were made, one captures especially well the normative views of education, combined with critical perspectives on AI, that were explored in this chapter (see Figure 23.5).

This placard quotes the conclusion of Camus's (2013) 1942 essay, *The Myth of Sisyphus*, to articulate a position for humanity when facing platforms that promise effortlessness in place of struggle. The placard makes clear the risk: of AI replacing students by eliminating their struggle. In this future, people's words, ideas, and identities might be 'fed to ourselves on a silver platter', removing the satisfaction that comes from hard work and challenge.

The form of 'success' that is evoked here initially looks like a fairly standard portrayal of the competitive, responsibilised individual success prominent in education in the early twenty-first century (Torrance, 2017). But the demand itself proposes something different. It frames the task of education as an endless struggle: not to 'finish on top', as the top will never be reached, but to continually struggle to learn.

Should we understand RAI as a partner in that struggle, and what would *that* look like? The potential for human flourishing in an age of AIED is seen, by young people, as feasible, but it requires educational values that are at odds with the efficiency drive being pursued by many EdTech and AIED proponents. This may be a challenging shift to make, as young people themselves struggled to see Responsible AIED in other-than-efficiency terms.



Source: Authors' own.

Figure 23.5 *The value of struggle, hard work, and happiness: 'We should work for success, not feed it to ourselves in a silver platter'*

Still, one of the things we found notable in our workshops was the value young people placed on AI's potential to support more equitable experiences of effort – where what might be demoralising struggle is minimised in favour of more productive struggle (Murdoch et al., 2020): 'in which one engages effortfully to understand something unfamiliar' (p. 660). Murdoch et al. (2020) focus on teacher-student relationships and conversations as a mediator of productive struggle, and contrast these with both unproductive and destructive struggle (p. 666). There is useful work to do here to explore these concepts further in the context of GenAI in educational settings. We recommend this as a future direction for critical AI and education research, and a possible source of hopeful engagement with young people around education futures in the context of continually emerging technologies.

CONCLUSION

This chapter explored the nature of young people's engagement with AIED futures, identifying both the critical perspectives young people can bring to discussions of GenAI and the need to proceed cautiously with fostering young people's critical AI literacies and perspectives. Such work requires critically *hopeful* framings of educational technology futures. It needs to avoid sparking resignation and cynicism about education amongst young people who need school to be a safe and trusted space in a complex socio-technical landscape where trusted institutions are in short supply.

This is not to suggest that the nature of the trust young people hold is either justified (see, e.g., Hope (2008) on Internet overblocking and its consequences) nor inevitable in its current form. The patrolling of boundaries in an attempt to enforce young people's safety and compliance brings issues of surveillance and monitoring into sharp relief, and works against the hybridity of learning contexts and capabilities that young people both experience and need (Nørgård, 2021). However, it does suggest a need to tread very carefully when attempting to mobilise young people's voices around RAI in education. Doing so risks *responsibilising* them inappropriately, and also negatively affecting their attitudes of trust towards education and schooling. If young people believe that schools can be a force for digital good and responsible technology practice in a datafied world, perhaps we can find more ways to help schools earn this faith in an age of Generative AI.

ACKNOWLEDGEMENTS

Our thanks to the UK's Arts and Humanities Research Council and the Bridging AI Responsible Divides programme. The research that underpins this chapter was funded by research grant AH/Z505560/1. Thanks to Dr Joe Noteboom, whose work on datafication in Higher Education informed our thinking during analysis for this chapter.

NOTES

1. A thought experiment based on scenarios where taking no action will result in harm to a number of people, while taking some action will harm just one, different person. https://en.wikipedia.org/wiki/Trolley_problem (retrieved May 2025).
2. It is possible that the in-built characteristic of AI tools, drawing on image banks populated by images from the past/present, may have hindered a more radical imagination of future education. For instance, AI tools produced images of 'future' schools populated with rows of desks and chairs, blackboards, clusters of mostly white children in school uniform, and Victorian school buildings/architecture. AI tools, as they currently exist, may not be the most useful companion for prompting speculative and alternative imaginaries.
3. Scottish advice service for additional support for learning: https://enquire.org.uk (retrieved October 2024).

REFERENCES

Andries, V., & Robertson, J. (2023). Alexa doesn't have that many feelings: Children's understanding of AI through interactions with smart speakers in their homes. *Computers and Education: Artificial Intelligence*, *5*, 100176. https://doi.org/10.1016/j.caeai.2023.100176

Antonenko, P., & Abramowitz, B. (2023). In-service teachers' (mis)conceptions of artificial intelligence in K-12 science education. *Journal of Research on Technology in Education*, *55*(1), 64–78. https://doi.org/10.1080/15391523.2022.2119450

Atabey, A., Botta, C., Dyer, H., Priyadharshini, E., Ross, J., & Wilson, C. (2024). *What's at stake? Young people's take on AI & education*. University of Edinburgh.

Auger, J. (2013). Speculative design: Crafting the speculation. *Digital Creativity*, *24*(1), 11–35. https://doi.org/10.1080/14626268.2013.767276

Bai, Z., Judd, F., Polinsky, N., & Yadollahi, E. (2023). *Participatory design of AI with children: Reflections on IDC design challenge*. *arXiv.* https://doi.org/10.48550/arXiv.2304.09091

Bewersdorff, A., Zhai, X., Roberts, J., & Nerdel, C. (2023). Myths, mis- and preconceptions of artificial intelligence: A review of the literature. *Computers and Education: Artificial Intelligence*, *4*(2023), 1–11. https://doi.org/10.1016/j.caeai.2023.100143

Blackman, R. (2022). *Ethical machines: Your concise guide to totally unbiased, transparent, and respectful AI*. Harvard Business Review Press.

Camus, A. (2013). *The myth of Sisyphus*. Penguin UK.

Coeckelbergh, M. (2020). Artificial Intelligence, responsibility attribution, and a relational justification of explainability. *Science and Engineering Ethics*, *26*(4), 2051–2068. https://doi.org/10.1007/s11948-019-00146-8

Dai, Y., Lin, Z., Liu, A., & Wang, W. (2024). An embodied, analogical and disruptive approach of AI pedagogy in upper elementary education: An experimental study. *British Journal of Educational Technology*, *55*(1), 417–434. https://doi.org/10.1111/bjet.13371

Decuypere, M., & Landri, P. (2021). Governing by visual shapes: University rankings, digital education platforms and cosmologies of higher education. *Critical Studies in Education*, *62*(1), 17–33. https://doi.org/10.1080/17508487.2020.1720760

Dignum, V. (2021). The role and challenges of education for responsible AI. *London Review of Education*, *19*(1), 1–11.

Domínguez Figaredo, D., & Stoyanovich, J. (2023). Responsible AI literacy: A stakeholder-first approach. *Big Data & Society*, *10*(2), 1–15. https://doi.org/10.1177/20539517231219958

Draper, N. A., & Turow, J. (2019). The corporate cultivation of digital resignation. *New Media & Society*, *21*(8), 1824–1839. https://doi.org/10.1177/1461444819833331

Dunne, A., & Raby, F. (2013). *Speculative everything: Design, fiction, and social dreaming*. The MIT Press. https://mitpress.mit.edu/9780262019842/speculative-everything/

European Commission. (2019). *Ethics guidelines for trustworthy AI*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Hargittai, E., & Marwick, A. (2016). 'What can I really do?' Explaining the privacy paradox with online apathy. *International Journal of Communication*, *10*(2016), Article 0. https://doi.org/10.5167/uzh-148157

Hay, K. (2022). Zineography: Exploring the participatory design process of collaborative zine making. In *Proceedings of the Participatory Design Conference 2022* (Vol. 2, pp. 313–316). https://doi.org/10.1145/3537797.3537866

Higgs, J. M., & Stornaiuolo, A. (2024). Being human in the age of Generative AI: Young people's ethical concerns about writing and living with machines. *Reading Research Quarterly*, *1*, 1–19. https://doi.org/10.1002/rrq.552

Hope, A. (2008). Internet pollution discourses, exclusionary practices and the 'culture of overblocking' within UK schools. *Technology, Pedagogy and Education*, *17*(2), 103–113. https://doi.org/10.1080/14759390802098599

Hossain, S., & Ahmed, S. I. (2021). Towards a new participatory approach for designing artificial intelligence and data-driven technologies. *arXiv*. https://doi.org/10.48550/arXiv.2104.04072

Jensen, L. X., Buhl, A., Sharma, A., & Bearman, M. (2024). Generative AI and higher education: A review of claims from the first months of ChatGPT. *Higher Education*. https://doi.org/10.1007/s10734-024-01265-3

Jurenka, I., Kunesch, M., McKee, K. R. et al. (2024). Towards responsible development of Generative AI for education: An evaluation-driven approach. *arXiv*. https://doi.org/10.48550/arXiv.2407.12687

Kearns, M., & Roth, A. (2019). *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press.

Kenny, S., & Antle, A. N. (2024). Reimagining AI: Exploring speculative design workshops for supporting BIPOC youth critical Ai literacies. *arXiv*. https://doi.org/10.48550/arXiv.2407.08740

Kim, K., Kwon, K., Ottenbreit-Leftwich, A., Bae, H., & Glazewski, K. (2023). Exploring middle school students' common naive conceptions of Artificial Intelligence concepts, and the evolution of these ideas. *Education and Information Technologies*, *28*(8), 9827–9854. https://doi.org/10.1007/s10639-023-11600-3

Lao, Y., & You, Y. (2024). Unraveling Generative AI in BBC News: Application, impact, literacy and governance. *Transforming Government: People, Process and Policy*. https://doi.org/10.1108/TG-01-2024-0022

Linderoth, C., Hultén, M., & Stenliden, L. (2024). Competing visions of artificial intelligence in education – a heuristic analysis on sociotechnical imaginaries and problematizations in policy guidelines. *Policy Futures in Education*, 1–17. https://doi.org/10.1177/14782103241228900

Liu, J., & Bray, M. (2022). Responsibilised parents and shadow education: Managing the precarious environment in China. *British Journal of Sociology of Education*, *43*(6), 878–897. https://doi.org/10.1080/01425692.2022.2072810

Lukens, J., & Disalvo, C. (2011). Speculative design and technological fluency. *International Journal of Learning and Media*, *3*, 23–40. https://doi.org/10.1162/IJLM_a_00080

Luo, J. (2024). How does GenAI affect trust in teacher-student relationships? Insights from students' assessment experiences. *Teaching in Higher Education*, 1–16. https://doi.org/10.1080/13562517.2024.2341005

Lyon, D., & Carabelli, G. (2016). Researching young people's orientations to the future: The methodological challenges of using arts practice. *Qualitative Research*, *16*(4), 430–445. https://doi.org/10.1177/1468794115587393

MacLure, M. (2013). The wonder of data. *Cultural Studies ↔ Critical Methodologies*, *13*(4), 228–232. https://doi.org/10.1177/1532708613487863

MacLure, M. (2023). Transversal inquiry: The 'adventure of the involuntary'. In L. A. Mazzei & A. Y. Jackson (Eds.), *Postfoundational approaches to qualitative inquiry*. Routledge.

Malinverni, L., Mora-Guiard, J., Padillo, V., Mairena, M., Hervás, A., & Pares, N. (2014). Participatory design strategies to enhance the creative contribution of children with special needs. In *Proceedings of the 2014 Conference on Interaction Design and Children* (pp. 85–94). https://doi.org/10.1145/2593968.2593981

McClain, C., Faverio, M., Anderson, M., & Park, E. (2023). *How Americans view data privacy: The role of technology companies, AI and regulation – plus personal experiences with data breaches, passwords, cybersecurity and privacy policies*. Pew Research Center. https://www.jstor.org/stable/resrep57319

McDonald, N., Johri, A., Ali, A., & Hingle, A. (2024). Generative Artificial Intelligence in Higher Education: Evidence from an analysis of institutional policies and guidelines. *arXiv*. https://doi.org/10.48550/arXiv.2402.01659

McGrath, C., Cerratto Pargman, T., Juth, N., & Palmgren, P. J. (2023). University teachers' perceptions of responsibility and artificial intelligence in higher education – an experimental philosophical study. *Computers and Education: Artificial Intelligence*, *4*, 1–9. https://doi.org/10.1016/j.caeai.2023.100139

Miao, F., & Holmes, W. (2023). *Guidance for Generative AI in education and research* [Report]. United Nations Educational, Scientific and Cultural Organization (UNESCO)). https://unesdoc.unesco.org/ark:/48223/pf0000386693

Microsoft. (2022, June). Microsoft Responsible AI Standard, v2. https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf

Murdoch, D., English, A. R., Hintz, A., & Tyson, K. (2020). Feeling heard: Inclusive education, transformative learning, and productive struggle. *Educational Theory*, *70*(5), 653–679. https://doi.org/10.1111/edth.12449

Nørgård, R. T. (2021). Theorising hybrid lifelong learning. *British Journal of Educational Technology*, *52*(4), 1709–1723. https://doi.org/10.1111/bjet.13121

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Onal, S., & Kulavuz-Onal, D. (2024). A cross-disciplinary examination of the instructional uses of ChatGPT in Higher Education. *Journal of Educational Technology Systems*, *52*(3), 301–324. https://doi.org/10.1177/00472395231196532

Pyysiäinen, J., Halpin, D., & Guilfoyle, A. (2017). Neoliberal governance and 'responsibilization' of agents: Reassessing the mechanisms of responsibility-shift in neoliberal discursive environments. *Distinktion: Journal of Social Theory*, *18*(2), 215–235. https://doi.org/10.1080/1600910X.2017.1331858

Regmi, K. D. (2023). Meritocratic lifelong learning: Responsibilisation of marginalised adults for their learning as neocolonial contract. *International Journal of Lifelong Education*, *42*(4), 406–423. https://doi.org/10.1080/02601370.2023.2231640

Ross, J. (2022). *Digital futures for learning: Speculative methods and pedagogies*. Routledge. https://doi.org/10.4324/9781003202134

Salinas, A., Shah, P. V., Huang, Y., McCormack, R., & Morstatter, F. (2024). The unequal opportunities of large language models: Revealing demographic bias through job recommendations. *arXiv*. https://doi.org/10.48550/arXiv.2308.02053

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education*, *57*(4), 620–631. https://doi.org/10.1111/ejed.12532

Shamir, R. (2008). The age of responsibilization: On market-embedded morality. *Economy and Society*, *37*(1), 1–19. https://doi.org/10.1080/03085140701760833

Sperling, K., Stenberg, C.-J., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In search of artificial intelligence (AI) literacy in teacher education: A scoping review. *Computers and Education Open*, *6*, 1–13. https://doi.org/10.1016/j.caeo.2024.100169

St. Pierre, E. A. (2021). Why post qualitative inquiry? *Qualitative Inquiry*, *27*(2), 163–166. https://doi.org/10.1177/1077800420931142

Su (苏嘉红), J., & Yang (杨伟鹏), W. (2023). Unlocking the power of ChatGPT: A framework for applying Generative AI in Education. *ECNU Review of Education*, *6*(3), 355–366. https://doi.org/10.1177/20965311231168423

Torrance, H. (2017). Blaming the victim: Assessment, examinations, and the responsibilisation of students and teachers in neo-liberal governance. *Discourse: Studies in the Cultural Politics of Education*, *38*(1), 83–96. https://doi.org/10.1080/01596306.2015.1104854

Torres, A. P. G. (2023). Responsible AI: Law and advancing moral responsibilization. Conference on Technology Ethics – Tethics, Turku, Finland, 147–155. https://ceur-ws.org/Vol-3582/FP_12.pdf

Van Den Berg, G. (2024). Generative AI and educators: Partnering in using open digital content for transforming education. *Open Praxis*, *16*(2), 130–141. https://doi.org/10.55982/openpraxis.16.2.640

Welbl, J., Glaese, A., Uesato, J., Dathathri, S., Mellor, J., Hendricks, L. A., Anderson, K., Kohli, P., Coppin, B., & Huang, P.-S. (2021). Challenges in detoxifying language models. *arXiv*. https://doi.org/10.48550/arXiv.2109.07445

Whitman, M. (2020). 'We called that a behavior': The making of institutional data. *Big Data & Society*, *7*(1), 1–13. https://doi.org/10.1177/2053951720932200

Whittaker, M., Alper, M., Bennett, C. L. et al. (2019, November 20). *Disability, bias, and AI – repor*t. *A*I Now Institute. https://ainowinstitute.org/publication/disabilitybiasai-2019

Yeh, H.-C. (2024). The synergy of Generative AI and inquiry-based learning: Transforming the landscape of English teaching and learning. *Interactive Learning Environments*, 1–15. https://doi.org/10.1080/10494820.2024.2335491

YoungScot. (2024). Young Scot Big Survey results. Young Scot Corporate. https://youngscot.net/shop/young-scot-big-survey-results