JODDE Journal of Open, Distance, and Digital Education

Dron, J. & George, S. (2026). A matter of principle: A response to Stephen Downes's "On ethical AI principles". *Journal of Open, Distance, and Digital Education, 3*(1), 1-20. https://doi.org/10.25619/129g6j48

Received: 07 April 2026 | Revised: 28 April 2026 | Accepted: 28 April 2026

# A matter of principle: A response to Stephen Downes's "On ethical AI principles"

Jon Dron[1], Stella George[1]

[1] Faculty of Science and Technology, Athabasca University, Athabasca, Canada

Correspondence:
Jon Dron | jond@athabascau.ca

Abstract

This paper is written as a response to Stephen Downes's "On ethical AI principles", that we recommend should be read in advance. Though framed as a critique, it builds on Downes's work, filling in gaps and arguments that we believe to be missing from the original and challenging some assumptions, while acknowledging the soundness of the underlying messages. Our central point of disagreement stems from Downes's implied definition of the word "principle". We extend the argument from Downes's generic concerns about principles to a more specific educational context, highlight fuzziness in definitions, and reframe Downes's concerns that principles are political as more general issues of power.

Keywords

Principle; AI; generative AI; GenAI; education; ethics; technology

# 1 Introduction

This paper came about following a review by the first author, Jon Dron, of the draft version of Stephen Downes's paper, On ethical AI principles (Downes, 2026). Downes politely declined to make most of the requested changes, not unreasonably claiming that it would essentially mean writing a new paper. The editors suggested Dron should instead write a riposte, to be published in the same issue. Conscious of his own limitations in both expertise and time, Dron sought help from the second author, Stella George. This paper is the result.

As we progressed, it became clear that we were as much writing the paper that we wished Downes had written as we were writing a riposte to it. In it, we build on his foundations, extend some arguments, and highlight many points of agreement, but we will also be making the following central criticisms:

- That developing sets of principles is seldom if ever meant to create immediate consensus or cohesion, nor to mandate their universality. The intent of principles is normally to provide a means of expressing, discussing, and operationalizing ethical opinions and assumptions within a specified context.
- That there are features of educational systems, most notably in the power relationships they embody with a purpose to change how people think, that raise distinctive ethical issues of interest to readers of this journal; Downes did not address them.
- That Downes overgeneralizes the term "AI" and thus treats GenAIs like any other tool. We observe that GenAIs present novel ethical challenges thanks to their capability of themselves using tools, from language to software, and we explore some of the ramifications.
- That Downes conflates power and politics. Everyone makes ethical decisions, but those made by the rich and politically influential (as well as others with power, such as teachers in a classroom or parents of children) have more impact.

# 2 Principles as ethical tools

The avowed purpose of Downes's paper (2026, p. 1) is to demonstrate that "there is … no set of ethical principles that can be regarded as a consensus opinion on the ethical use of artificial intelligence in education". We believe that this sets an extremely low bar because, to the best of our knowledge, that is almost never their intent.

Downes does not tell us how he defines "principle", instead asking Google's search AI [sic] for some examples. As well as leaving what "principle" means open to Google's search AI's interpretation, the effect of doing so is to decontextualize the principle from its application. We think this matters because, for the most part, principles can, should and do, mean different things to different people in different contexts. The principle of not causing physical harm to others, for example, applies differently in a boxing match, where only some forms of harm are prohibited, than when applied to a caregiver in a kindergarten, a fact that is fully understood by all participants in each context: the meaning is, in each case, that which is understood by the specific community in which it is used. Downes provides plenty of similar examples.

We believe, following both dictionary definitions and custom and practice, that principles are foundational guidelines, starting points, and orientations that are used to frame understanding and assist with decisions. Like Dworkin (1967), we see principles as intimately connected with but distinct from rules. Principles are what we turn to when rules are ambiguous, inapplicable, inadequate, or missing. This is, for example, how they are used in the the European General Data Protection Regulations, as guides to interpretation and intent (*General Data Protection Regulation*

*(GDPR) Compliance Guidelines*, 2018). We believe that what Downes is actually arguing against are in fact better understood as laws, that must both be followed and that are considered to be universal. A few ethical frameworks do attempt to use "principle" in this way. Kant's categorical imperative, for instance, a principle in its own right, is explicit that a maxim (which could also loosely be interpreted as a principle) of an action should be universally applicable, if only hypothetically (Kant, 1925/2017). However, it results in the same conundrums that any ethical framework encounters when faced with the simple fact that principles can conflict, and solutions must take context into account. Also, it is not how (at least) many people actually make ethical decisions. In real life, amongst other things, people are affected by moral contagion, have diverse moral compasses, bring conflicting moral dimensions (effectively principles, such as doing no harm, being honest, etc) to bear, and are influenced by political or religious beliefs, as well as mood and pragmatic concerns (Hofmann et al., 2014).

It is not only possible but very common for principles to be in conflict, and ethical decision-making is largely concerned with resolving such conflicts. A conflict between principles of avoiding harm to others and telling the truth, say, might occur in many situations from negotiating the release of hostages to being polite about a bad haircut. Principles can even conflict with themselves: it is common to choose the lesser of two harms, for example, as the Trolley Problem (Thomson, 1984) well illustrates. By identifying the principles involved, we are better able to reflect on, debate, refine, and come to more reasoned ethical decisions. Principles are not the means through which we resolve such ethical dilemmas, but through which we articulate the issues and rehearse, establish, or negotiate our positions. There are other ways this might be accomplished, including Downes's sentiment-based moral particularism that he describes in his conclusion, and there are certainly more sophisticated debates needed in order to identify appropriate principles in the first place, but principles are popular, easily communicated, effective tools that scale well and, importantly, act as boundary objects (Star & Griesemer, 1989) between different cultures and communities. Their boundary-object role does not make them universalizable: in fact, it draws attention to their negotiable and flexible meanings as contexts shift. Like any tool, they can serve other purposes, and they are useless without further orchestration. This may well include other ways of doing ethics – consequentialist, deontological, virtue-based, particularist, or whatever – but the principles are essential parts of the whole. We believe, like Aristoteles et al. (1925/1991, p. 2), that ethics is concerned with things that are *only for the most part true*. We share with Downes a concern that principles are often ambiguous, easily misconstrued, that they can be bent to political ends, or used to inform poorly considered legislation. However, principles are tools, not ends in themselves. We would be considerably more concerned if politicians and legislators did not have any principles.

Principles are part of the fabric of most educational institutions, often appearing in or informing mission statements, policies, and procedures, where they are frequently used to express the distinctiveness of the culture and practices of the institution. While most are blandly interchangeable, the fact that each institution does at least try to have its own speaks to their non-universality, and that the differences between them are (again) perceived as a feature, not a bug.

This in turn speaks to the fact that principles do not stand alone: they exist in relation to a myriad of others and are parts of a broader network of values, the weights of which constantly shift according to context, preference, and the relative weights of their neighbours. The same principle may justify two very different but equally defensible courses of action, but it is its configuration in relation to the rest of our values that will determine which we choose.

Following from this, we do not think that Downes's use of a GenAI to provide a list of examples is a bad decision, *per se*, as long as the principles reasonably accurately mirror those actually used by humans in the context in question. If the intent is to argue against a representative and typical

sample of principles then all that really matters is that the reader agrees that this is what they are and, given that the large language models (LLMs) behind GenAIs like Google's search AI are the result of training on vast datasets, it is likely that they will provide an approximating-to-average (albeit a culturally biased average) summary of typical principles. However, these principles all emerged in specific contexts that are lost in Google's search AI's reformulation of them. Without context, their meaning and relative significance become open to interpretation.

Downes does not tell us the precise search term that he used, but his description of asking for "major principles of ethical AI" and our own attempts to replicate this suggest that he did not limit it to the field of education. This leads us to the second major issue that we find with the paper: that it largely ignores the context.

## 3 The ethics of AI in education

In his abstract, Downes (2026, p. 1) claims: "Some argue that the ethical use of artificial intelligence in education can be defined through the use of a set of commonly held principles, such as sustainability, accountability, and fairness. This article challenges that presumption."

However, none of Downes's principles appear to have been drawn from an educational context, and a large majority of his arguments explicitly refer to almost anything but. The only significant discussion of the educational context occurs in one short section, the purpose of which is almost entirely scene-setting. This matters because we believe that there are some distinctive concerns that distinguish the use of AI in education from its broader uses, including:

- The power imbalance of the teacher/student relationship. This asymmetry makes it possible, for instance, for teachers either to require or to prohibit the use of AI, and to be quite specific about what is and is not appropriate or expected. Equally, teachers have an explicit duty of care that is uncommon in most occupations.
- The formative nature of the system. Educational systems are designed to change who you are. Amongst other things, this is relevant to uses of AI when, even slightly, we learn to be human from things that are not quite human. It is particularly concerning that we may delegate some of the "thinking" part of a task to an AI, especially when that involves creative thought, because it is precisely that creativity that we may wish to nurture (Fernandes et al., 2026; Gerlich, 2025). Concerns have also been expressed about the risks of cognitive decline posed by GenAIs (Barcaui, 2025; Kosmyna et al., 2025). While we believe such studies shed more light on the weaknesses of the tasks students are expected to perform than any fundamental problem with GenAI, this is all the more reason to highlight it as a matter of ethical concern, if only because of the potential positive benefits of evolving away from the medieval or industrialized pedagogies that prevail in our schools and other educational institutions.
- The fact that time spent interacting with an AI may eat into time we might otherwise spend interacting with humans means that there are risks of not learning to work with other humans quite as effectively, or that the tireless sycophancy of typical GenAI chatbots may inculcate unwanted habits that are unhealthily transferred to human relationships (Cheng et al., 2026; Fang et al., 2025).
- The relational nature of the process. Educational systems are fundamentally rooted in human relationships (Bingham & Sidorkin, 2004; Buber, 1947). This is particularly significant in the context of AI chatbots, which play human-like roles and that may, in some ways, be considered to be partners in the learning process. Unfortunately, current generations of chatbots are neither genuinely autonomous nor are they persistent: their

identities are fluid and can change in an instant with a single prompt. They are thus not moral agents in the same sense as humans.

A further complication arises from the potential of (especially generative) AIs of having harm done to them by being trained on datasets created by GenAIs: so-called "model collapse" (Shumailov et al., 2024). While, at least in their current forms, it is not reasonable to ascribe moral agency to an AI nor any intrinsic moral worth of its own, the large-scale systemic impacts as they become an increasingly essential part of our educational and societal infrastructure are potentially, slowly, and subtly catastrophic. There are meaningful ethical questions to be asked about the whole-system consequences of AIs's potential dumbing down and the educational implications. For instance, do educators have a duty to counter this risk in not just pedagogical but curricular choices?

To help illustrate the significance of this context, in Table 1 we present comparisons between Downes's principles, examples of general principles that have actually been proposed by various human researchers and organizations, and those that have specifically been created by humans for use in the context of education. These are gleaned from a meta-analysis of a convenience sample of academic and significant organizational sources (Adams et al., 2023; China Association for International Science and Technology Cooperation, 2017; General Data Protection Regulation (GDPR) Compliance Guidelines, 2018; Holmes et al., 2022; Jobin et al., 2019; Pardo & Siemens, 2014; Prinsloo & Slade, 2017; Selwyn, 2019; Slade & Prinsloo, 2013; UNESCO, 2021; Villani et al., 2018), from which we collected and abstracted sets of principles. Doubtless, a different sample would lead to slightly different findings, but it is exactly our point that principles are situated in contexts, so we think this is sufficient to demonstrate the effect.

From our analysis, it can be seen that, as expected, the principles generated by the GenAI do broadly (if imperfectly) align with a representative sample of generic principles generated by humans. However, there are significant differences, refinements, and augmentations when applied to education, with 9 of the 20 principles *only* showing up in educational literature and specific educational issues emerging even for those that are common. Different contexts would lead to different lists. For example, if we were considering AI use in online adult education, a relevant principle might address the risks of increasing already problematic social isolation, or perhaps the ergonomic impacts of greater computer use. Again, this speaks to the non-universality of principles and their value in providing foci for dialogue. They do not tell us what to do, but instead are used to build and resolve ethical dilemmas in the contexts in which they arise.

By allowing his principles to apply across any and all contexts, Downes is able to construct apparent contradictions that disappear when context is taken into account. For example, to borrow one of his rare examples from education, cannibalism *would* almost certainly be presented fairly as a lifestyle choice in a society in which it was an acceptable practice.

Table 1: A comparison between Downes's AI-generated principles, human-created principles, and human-created principles in the context of education. (Legend: core indicates literature position named in all comparators; common present in some form in all comparators; peripheral named in two of three comparators; education specific, related to AIED literature only).

| | Principle | Positioning in literature | Downes's principles | General AI ethics literature | AI ethics in education (AIED) |
|---|---|---|---|---|---|
| 1 | Fairness and Equity | Core | Fairness: AI systems should not discriminate against individuals or groups. This includes ensuring that training data is representative and that models are not biased. | Near-universal (Jobin et al., 2019); includes bias mitigation, equitable distribution of benefits and burdens. | Goes beyond statistical bias to structural educational inequality; AIED may reproduce systemic disadvantage. |
| 2 | Transparency and Explainability | Core | Transparency and Explainability: AI systems should be designed in a way that allows users to understand how they work. This can help build trust and identify potential biases or errors. | Near-universal; top-ranked principle across 84 guidelines (Jobin et al., 2019); UNESCO first global standard; included in GDPR (2018). | Applied specifically to assessment decisions, predictive risk scores, and automated feedback that can harm educational trajectories. |
| 3 | Accountability and Responsibility | Core | Accountability: There should be clear lines of responsibility for the development, deployment, and use of AI systems. This helps ensure that AI is used responsibly and that those responsible can be held accountable for any negative consequences. | Near-universal; clear human responsibility for AI behaviour; EU, UK, UNESCO all emphasise this. Some literature refers to lawfulness. | Raises specific question: who is responsible when an AI tutoring system embeds poor pedagogy, or mislabels a student as at-risk? |
| 4 | Non-maleficence | Core | Non-maleficence: AI systems should be designed to avoid harming individuals, society, or the environment. This includes considering the potential risks and consequences of AI systems before they are deployed. | Near-universal (Jobin et al., 2019); covers unsafe design, misuse, unintended consequences. | Intersects with obligation to act on learning data — if student data shows risk, non-maleficence may require intervention (Prinsloo & Slade, 2017). |

| | Principle | Positioning in literature | Downes's principles | General AI ethics literature | AI ethics in education (AIED) |
|---|---|---|---|---|---|
| 5 | Privacy and Data Protection | Common | Privacy and Data Protection: AI systems should protect individuals' personal data and respect their privacy, ensuring that data is collected and used ethically and that individuals have control over their data. | Top-five across all guidelines considered covering: UNESCO, EU, UK, emphasise protection throughout AI lifecycle (Jobin et al., 2019) and via GDPR (2018); and in China (CAISTC*, 2017). | Significantly inflected Learning analytics adds: power asymmetry (students cannot freely withdraw), developmental sensitivity of children's data, informed consent challenges. |
| 6 | Inclusiveness | Common | Inclusiveness: AI systems should be designed to be accessible to all users, regardless of their background or abilities. This can help ensure that AI is used to benefit all members of society. | Present but less consistently articulated; related to solidarity and fairness in general frameworks. Strong represented for France, Villari (2018). | Inflected. Must not reproduce the digital divide; must accommodate neurodiversity; must serve globally diverse student populations rather than encoding Western assumptions. |
| 7 | Reliability and Safety | Common | Reliability and Safety: AI systems should be reliable and safe to use. Includes ensuring that they are not vulnerable to errors or malicious attacks. | Present in UK and at UNESCO (Jobin et al., 2019), CAISTC* (2017), USA (Trump, 2019, Biden, 2023). Nations commonly position AI safety in their standards. | Inflected. Includes reliability of assessment outputs; safety from algorithmic misclassification in high-stakes educational decisions. |
| 8 | Human-centered design | Common | Human-centered design: AI systems should be designed with the needs of humans in mind. AI systems are easy to use, understandable, and beneficial. | Related to beneficence and human autonomy; UNESCO (2021) emphasises human oversight, this is widely cited. | Inflected. In education must also attend to learner agency, developmental stage, and whether systems support or undermine self-directed learning. |
| 9 | Sustainability | Peripheral | Sustainability: AI systems should be sustainable and environmentally responsible. Includes considering the energy consumption and other environmental impacts of AI. | Present in extended Jobin et al. (2019) list; UNESCO (2021) links to UN Sustainable Development Goals. | Not distinctively developed or a focus of the AIED-specific literature; addressed through general frameworks when referenced at all. |

| | Principle | Positioning in literature | Downes's principles | General AI ethics literature | AI ethics in education (AIED) |
|---|---|---|---|---|---|
| 10 | Beneficence | Common | Not explicitly covered, however, may be considered to be subsumed under human-centred design and non-maleficence. | AI should actively promote human wellbeing, distinct from non-maleficence (Jobin et al., 2019). | Obligation to act on learning data (Prinsloo & Slade, 2017), a duty to intervene itself requiring ethical framing to ensure beneficence. Related to human-centred design (principle 8, above in this list). |
| 11 | Human autonomy and freedom | Common | Not explicitly mentioned as a standalone principle but perhaps implied. | AI should preserve meaningful human agency and freedom of choice (Jobin et al., 2019). Can also be refered to as social ethics. | Learner autonomy acquires a developmental and pedagogical character (Selwyn, 2019): concern that surveillance and datafication limit individual choices; stakes are development of self-directed, critical learners. |
| 12 | Human dignity | Peripheral | Not covered by selection. | Present in Jobin et al. (2019) extended list; AI must respect the inherent worth of persons. | Addressed implicitly; not a foregrounded principle in the AIED-specific literature. |
| 13 | Solidarity | Peripheral | Not covered by selection. | AI should serve common goods; not deepen social fragmentation (Jobin et al., 2019). | Related to inclusion and equitable access in AIED, but solidarity as such is not a named principle. |
| 14 | Pedagogical appropriateness | Education specific | Not covered by selection. | ABSENT No direct equivalent in general AI ethics frameworks. | Adams et al. (2023) asks whether AI tools embody sound theories of learning, rather than reducing education to what is easiest to model. No analogue in general frameworks. |
| 15 | Children's rights and developmental sensitivity | Education specific | Not covered by selection. | ABSENT No direct equivalent in general AI ethics frameworks. | Unique characteristics and vulnerability of minors; age-appropriate, developmentally sensitive AI design. (Adams et al., 2023). Absent from general frameworks entirely. |

| | Principle | Positioning in literature | Downes's principles | General AI ethics literature | AI ethics in education (AIED) |
|---|---|---|---|---|---|
| 16 | Teacher agency and professional integrity | Education specific | Not covered by selection. | ABSENT No direct equivalent in general AI ethics frameworks. | Teacher well-being, workload, pedagogical relationships; concern about professional deskilling and erosion of teacher judgment by algorithmic systems (Adams et al., 2023). |
| 17 | AI literacy and critical engagement | Education specific | Not covered by selection. | No direct equivalent in general AI ethics framework. Training for labour force and popularisation is included in China's plan (CAISTC*, 2017) | Students and educators should be equipped to understand and critically evaluate AI (Adams et al., 2023), suggesting a simultaneous goal and safeguard. |
| 18 | Ethics of pedagogy itself | Education specific | Not covered by selection. | ABSENT No direct equivalent in general AI ethics frameworks. | AIED ethics must address not only how AI is used, but whether the pedagogical models embedded in AI systems are themselves ethically defensible (Holmes & Porayska-Pomsta, 2022). |
| 19 | Obligation to act on learning data | Education specific | Not covered by selection. | ABSENT No direct equivalent in general AI ethics frameworks. | If data reveals a student is at risk, institutions may have an obligation to intervene (Prinsloo & Slade, 2017) — raising tensions between autonomy, privacy, beneficence, and paternalism. |
| 20 | Epistemic justice | Education specific | Not covered by selection. | ABSENT No direct equivalent in general AI ethics frameworks. | Emerging. Whose knowledge systems and ways of knowing are encoded in — or marginalised by — AI systems? Particularly salient for indigenous and globally diverse student populations. |

* CAIST: (China Association for International Science and Technology Cooperation, 2017).

## 4 The nature of AI and intelligence

In the Abstract, Downes (2026, p. 1) provides an ostensive definition for "AI": " a set of digital tools that can perform functions traditionally limited to human capability, for example, reviewing, summarizing, translating, and composing."

This definition would exclude a full simulation of the connectome of a fruit fly (Shiu et al., 2024), while determining a mechanical calculator to be artificially intelligent. Some of Downes's own examples are of a similar nature. Had the tools been available Skinner would no doubt have used AI for his teaching machine, but it is stretching the definition to describe what he actually created as in any way intelligent. The same is true of some forms of learning analytics that, in our opinion, are no more intelligent than the simple aggregation, averaging, and clustering found in a typical database or spreadsheet. This speaks to the fuzziness of his paper about intelligence in general. For the most part it is treated as a given but, if there's no agreement what "intelligence" means, then it is difficult to come to any agreement about artificial intelligence's ethical uses. We observe that Downes has elsewhere written of the closely related concept of knowledge as "constituted of the sets of connections between entities, such that a change in one entity may result in a change in the other entity, and that learning is the growth, development, modification or strengthening of those connections." (Downes, 2022). Downes could thus make the case that calculators or behaviourist teaching machines are both artificial and at least parts of an intelligence and, though the same is true of doorknobs and paper, we would agree.

Without falling into the debates around intelligence in the context of narrow and general AI, recognising that human intelligence is multi-faceted and is oversimplified when defined by its individual dimensions, we acknowledge that advanced GenAIs appear to converse and create in remarkably human-like ways. We think that this is because it actually is, to a large extent, human. We believe that what makes human intelligence particularly special is that it is a fundamentally social phenomenon, where we share in the collective cognition of our species through tokens, symbols, objects, movements, and activities that together provide not just most of the grist but, at least in part, the mill of thought (Heyes, 2018). Importantly, this perspective acknowledges that intelligence is, in humans, not just intrasomatic. Rather, it is distributed through our tools, media, and other technologies, especially through language. There is a strong case to be made that almost all the intelligence GenAIs possess lies in the human-generated content on which they are trained. Though still artificial – represented in human-created artifacts or tokens – and not to belittle the algorithmic ingenuity involved in its enactment, the intelligence of an LLM, say, is thus primarily human.

## 5 The nature of GenAI

Downes tends to use "AI" as a catch-all term while making arguments that sometimes apply only to learning analytics, sometimes only to large language models, and only rarely to other implementations. This lack of definition has little impact on the arguments he makes, which are largely against ethical principles, whatever their provenance.

We think, though, that there is something distinctive about GenAI, in particular, that demands a slightly different ethical stance: that it is the first family of technologies ever invented that are themselves technology users and that, in agentic form, they can do so quite "independently." We use the scare quotes around "independently" in order to make it clear that we do not equate this with human independence. The current generation of mainstream genAI does not have a fixed, continuous, or persistent identity nor any identifiable self, but, if permitted, can make decisions that can have an ethical impact. In extremis, as the CEO of Palantir recently put it, 'Our product

is used, on occasion, to kill people' (Thomas, 2026). In an education system, the potential for harm is also high, from risks of coarsening cognition to children learning ways of being human from machines (Dron, 2023b). We would not go so far (yet) as to claim that GenAI can be ethical agents with rights of their own. What differentiates GenAI from all past technologies is that what it creates – the consequences of its use of other technologies – is neither accidental nor the direct result of intentional orchestration by human beings. Even if it were possible to meaningfully audit its "thinking" processes, the fact that it can draw on indefinitely many other technologies (including words, code, imagery, video, and, increasingly, physical objects in the world), assembling and orchestrating them in countless new and original ways, makes it inherently unpredictable. We are thus challenged with a new kind of unpredictable ethical actor: not conscious, not fully independent, not thinking as we do, not capable of being mistreated, but actively and creatively changing the world. It is understandable that people fear and feel a need to put boundaries around this unknown, and to establish rules to govern it. We share Downes's implied concerns that doing so before we fully understand the nature of the beast is more likely to backfire than not, and that any such rules will more likely reflect existing biases, power structures and norms than a reasoned understanding of its distinctive nature.

As long as we think of GenAI as "just a tool" then such efforts are unlikely to be effective or appropriate. Like almost any technology, a full description of the technologies of AI is inseparable from not just what they are used for but also the uses themselves: a full description of the technology must thus include the human technique and orchestration that goes into the assembly (Dron, 2023a). It is in that technique, as much as in the choices about what it might be used for, that the ethical dimension lies. While there is little to distinguish the ethics of conventional procedural approaches to AI from any other technology, GenAIs can be co-orchestrators of the technology of education, coparticipants in its enactment, capable of acting creatively (in the sense of doing what has not been exactly done before) with other technologies in ways that, in a limited way, closely resemble those of humans. Using such a technology differs in kind from any other, except perhaps those employing animals (without the moral obligation), because the results are neither deterministic nor random: instead, they are unprestatable. We cannot in principle or practice know in advance precisely how they will behave. One might argue that the same is true of any complex technology. The difference, though, is that GenAIs are *designed* to work this way. It is a necessary consequence of their ability to do what they do (Kalai et al., 2025). We can try to contain the danger through the use of counter-technologies, much as when we reduce the danger of stoves with extraction fans, but, almost by definition, the counter-technologies we can build will not be as smart – in the sense of being capable of adapting responses to novel situations – as those they attempt to contain. This is new ethical territory.

## 6 Power and politics

A recurring theme throughout Downes (2026, p. 1) is again introduced in the Abstract: "It is argued that much of what is offered as a set of ethical considerations reflects, in fact, a political argument and perspective."

Though framed as political, in his arguments themselves, Downes's target appears to really be power, including political power but especially that which is exercised by the richer and more privileged segment of the world's population. We think this could equally apply at a micro or meso scale to, say, a group of friends, a family, or a school.

Everyone makes ethical choices, but those of the more powerful are generally more impactful. It is true that ethical principles can privilege the views of a minority, serve selfish ends, or be used as instruments of dominance. In a very literal way, many of the ethical (or otherwise) choices

made by the likes of Elon Musk, Bill Gates, the Pope, the President of the USA, or Mark Zuckerberg impact far more people than the choices made by most children, or most women, or almost anyone in the global South. However, they are still the same kinds of ethical choice that everyone else makes. Law-makers and policy-makers can have a similarly disproportionate influence, as can religions (Friedman, 2007; Marx, 1844/1994). Colonialists have weaponized principles to justify the subjugation and abuse of indigenous populations (Mehta, 1999). It is power and influence, not politics, that makes this so, and it plays out at every scale. This is nowhere more obvious than in the power imbalances within educational systems themselves, and the potential harms that they make possible when combined with AIs, from analytics tools to chatbots to career advice applications. Like Downes, we are particularly concerned with ways that adaptive systems can be used to "personalize" learning (Downes, 2016; Kohn, 2015), especially when the targets are hard learning outcomes determined by someone else, especially when done at scale.

It is, however, important to distinguish between the ethical principles and the ways they are used, for they are tools that can be bent to many different purposes, in assembly with other technologies and other phenomena. This is precisely why proponents see them as having value in assembly with AIs: they can act as counter-technologies to the harms that, unbridled, people might do with AIs or AIs might do unaided. Quite apart from their increasing use to cause direct harm, there are sufficient examples of suicides, dangerous behaviours, and misinformation that are attributable to individuals' interactions with AIs, including those shaping our social media, to make this a matter of serious concern.

A great deal of effort goes into training LLMs to avoid what their trainers perceive as unethical behaviour and that does indeed mean that they can impose whatever ethical attitudes and even rules that they wish to impose. The scale of the impact of a small number of large genAI systems, globally, far from the original context in which the principles themselves evolved is concerning. This is a power problem: with, pragmatically, limited choice apart from to use one or more of these models, it imposes the tastes of a few on the many, whoever the few may be. It is, for instance, disturbing that educators are increasingly being sold on the idea of LLMs playing tutoring roles that may work well to teach the intended outcomes, but may not be sensitive to local needs and ethical norms, and that may teach a hidden/tacit curriculum that is inappropriate to a given community (Warr & Heath, 2025). This is not a new phenomenon, but the scale is different. The manifold harms that uses of AI in social media such as Meta's Facebook have caused or catalyzed, from political shifts, to the rise and growing dominance of ignorance-culture, serve as a sobering warning and prelude for what is happening now in the field of (in particular) generative and agentic AI. Without a means to articulate how ethical choices are made, we cannot begin to even talk about this, let alone make informed choices. Principles provide that framing.

# 7 Downes's arguments in detail

With the foregoing in mind, we turn to Downes's (2026) arguments against the principles themselves.

## 7.1 Fairness

Downes (2026) rightly observes that there is no single concept of fairness that applies across all contexts, presenting some entertaining ad absurdum arguments about what happens if the wrong one is applied in the wrong context. We think that he oversteps the bounds here and there, inasmuch as fairness is only an ethical issue when it is the result of actions of an ethical decision-maker. The happenstance of where you are born, for instance, is not an issue of fairness unless

you believe that some kind of deity is responsible for it. Similarly, to suggest that life isn't fair in anything other than a metaphorical sense is no more reasonable than claiming that life likes the colour pink, because life is not moral agent. However, we agree that what is fair in an Olympic race is not the same as what is fair in the allocation of candy. We agree, too, that the creator, owner, or enforcer of a principle making use of the concept is the one with the most power to set the boundaries of ethical decisions relating to it. None of this has anything to do with whether or not it is a moral principle. Bias causes harm to some individuals, so, in principle, it is fair to claim that it should be avoided. The fact that the principle may conflict with others, or be interpreted differently in different contexts, or be inequitably shaped or abused by those with greater power, does not detract from its innate ethical import. Giving it a name opens up discussions about exactly those concerns. It is especially important in the context of education, where minds are being formed, often with regional or national curricula that can have a vast impact on the future of a society. That every culture, subculture, and society may have a different understanding of what is and is not fair in any given context is immaterial to the question of whether it is an ethical principle.

### 7.2 Transparency and explainability

It is true that LLMs are not currently explicable with any precision, but there are plenty of AIs that are. Downes (2026) also assumes that the principle of explainability demands that an AI should be explainable to anyone and everyone, but his own final argument demonstrates why this makes no sense. It would be immoral of a manufacturer of sausages to purvey bratwursts containing excess toxins, bacteria, or non-foodstuffs not advertised on the label, but no one expects to have to carry a microscope or chromatography kit on each trip to their local bratwurst vendor: this is why we have food inspection agencies. Similarly, as long as experts are able to explain an AI's actions, it does not matter that we cannot all do so.

We do agree, however, that explainability is not itself a moral principle. It is a means of operationalizing the more basic principle of doing no harm. If the makers of a technology cannot know whether it will cause harm, and if there is a good chance that harm will be caused, then there are moral concerns about providing it to someone else. This is one of the reasons that suppliers of street drugs who cut them with other unrevealed and often more harmful substances are immoral while big pharmaceutical companies are (in this one respect) not. The disclaimers provided by producers of GenAIs are akin to drug dealers expressing partial ignorance of what they are selling: better than nothing, but not enough. Does this make explainability an ethical principle for all? We think not: it is, as Downes says, what makes it possible to enforce a principle, but it is a means, not an end. Popular GenAIs are appliances that can be used for indefinitely many purposes, and their developers cannot be expected to anticipate more than a small fraction of these.

### 7.3 Accountability

Downes (2026) points out, and we agree, that accountability is not an ethical principle in itself but part of the mechanism for enforcing principles related to responsible use such as non-maleficence and beneficence, fairness and equity, and also that there is nothing distinctive about AI in this. He then, however, spends some time arguing that AIs are not doing anything humans cannot do: on scale of action alone, we disagree. We think this misses the point, however. It is not AIs but their makers, vendors, and users who are accountable for what they do, just as Downes is accountable for the principles discussed in this paper even though Google's search AI was their producer. There are further complexities inasmuch as, at least in the use of chatbots and agents, the user may play a significant role in the "thinking" of the machine: their prompts are a literal part of it. Also, the capacity of GenAIs to generate novel results complicates things further

because, no matter how well tested a system might be, there are unprestatably many possible creations that could cause harm in unforeseeable ways. In this sense, a GenAI is closer in moral character to a pet dog. We do expect those who are responsible for it to take reasonable measures to prevent it biting or harassing others, and we do hold them accountable when it does do harm, but we recognize that, sometimes, despite that, dogs will do their own thing.

Downes (2026) wraps the section up with a more tangential observation that, in many cases, accountability is diffuse and distributed, and very unevenly applied. The fact that the principle can be used, bent, or ignored for political reasons and/or for greed does not make it non-ethical, nor inherently political, however. We have to know or, at least, be reasonably expected to know that something is unethical in order to be held responsible for unethical behaviour, and a principle is one way of communicating that.

### 7.4 Privacy and data protection

Downes (2026) makes three main arguments against privacy and data protection being thought of as ethical principles. The first (and last) is that there are many exceptions. We think we have dealt with this already. The fact that other principles may conflict with or override this principle, if anything, amplifies the point that it is a principle. There are indeed many reasons that some data should be protected more than other data, and some data, especially when harm to others is involved, that should never be private. Principles of data protection are concerned with identifying when and to what extent data should be protected, and what levels of control should be afforded to individuals over what data about them others are allowed to keep. They are not concerned with making everything private. Again, Downes mistakes a principle for a law.

The second point Downes (2026) makes is that there is nothing special about AI and privacy: that existing principles still apply. We agree, to a point, though we note that the scale and depth of privacy invasion that is possible with some forms of AI do make it a different kind of problem, just as the ability of machines to store and manipulate vast amounts of data have made privacy legislation necessary since the 1980s, or the destructive potential of guns compared with wooden sticks demands greater controls over their ownership and use. But, whether or not it is different, the question remains whether it is a purely political concern, or an ethical concern that may be utilized in a political context. We believe it is closer to the latter. However, like Downes, we are not convinced that it is an ethical principle in itself.

Privacy as a human right is, to a large extent, a modern invention (Friedman, 2007): for most of our history, there was very little of it and, if it were invaded, the problem was not the intrusion itself but the reason that the intrusion occurred, be that theft, shaming, discovery of plans, or whatever. It has become a concern as the means to invade it have become far more prolific and sophisticated, starting with the invention of photography in the 19th Century. However, in all cases that we can think of, the reason we need such a principle is not that privacy is an ethical good in itself, nor that the loss of privacy is necessarily bad, but that breaches of privacy can be used to cause further harm. That harm seems to us to be a better target of an ethical principle. There may be a need for legislation to protect us against misuses of privacy-invading technologies, just as there is a need for legislation to protect us against misuses of guns, but we agree with Downes that this is not an ethical principle in its own right.

The third point Downes (2026) makes appears to be that individuals or groups with power get away with invasions of privacy that the rest of us have historically been denied so we should be allowed to do it too. We assume this novel moral argument is intended to show the principle of privacy to be non-universal. As we have already argued, however, this is a straw man.

### 7.5 Inclusiveness

Downes (2026) initially takes inclusiveness as being mainly an issue of accessibility before leaving that behind and arguing that it is more to do with inequalities in wealth and resources. His arguments are, by now, familiar: that the principle is not universally upheld, that there are competing principles that may negate it, that it is unevenly distributed, and that there are many examples of individuals, cultures and societies that actively reject it. We refer the reader to our previous comments on why this is problematic. Downes does not argue that this is not an ethical issue so it is hard to know whether he believes it to be or not. We think it is, though we agree that there is nothing distinctive to AI or education in this: any technology that we require students to use should be usable by as many people as possible. To discriminate against a set of people who share common attributes is, on the whole, to do them harm.

### 7.6 Reliability and safety

Downes does a good job of explicating the complexities of a principle of reliability and safety. Indeed, as he puts it, "The principle, at best, can be expressed as 'all else being equal, we want AI to be reliable, safe and secure.'" (Downes, 2026, p. 10). We agree. All security is a compromise: a fully secure system would be virtually unusable for almost any task. We should, however, expect that providers, implementors, and users of AIs should, between them, take precautions that are proportionate to the risks in designing and testing AI products because, otherwise, they would knowingly cause harm. This is very much an ethical issue in education, especially in K-12 schools, where our power gives us an amplified duty of care. Downes does not use this section to further his argument: his main complaint is that those calling for it tend to be neo-Luddites. We agree, though this does not negate the fact that reliability and safety are ethically motivated requirements, nor that it is more complex to achieve this (at least for a GenAI) than it is for most technologies thanks to the difficulties predicting what it might do. By highlighting it as a principle we send the message that it needs special consideration, not that every system should be perfectly safe or secure.

### 7.7 Human-centred design

We agree with Downes (2026) that it is hard to see human-centred design as an ethical principle, except insofar as we should make the lives of others as comfortable as we can, all else being equal. Downes makes some interesting and acute observations in this section – we particularly like his comments on the mutual shaping of tools and their users - but the argument is let down by an apparent misconstrual of what "human-centred design" normally means. Downes repeatedly conflates "designed with human needs in mind" with "designed for humans to be in control of it," which is an important principle on occasions where humans should be in control (and a good part of what "human-centred" means in such a situation) but, as his examples demonstrate, is highly dependent on context. His final point, that putting humans in control keeps those who are in control in control is equally context dependent. For AI systems, human-in-the-loop is conceived to provide oversight to detect errors and drift in algorithm results, checking system veracity, not human-centred design. Anyone who can remember the dawn of the Web and the brief but welcome period when power shifted away from publishers and large media providers will be familiar with ways that technologies can be liberative and challenge dominant hegemonies. That "free" Web does still exist – indeed, it is thriving – though powerful predators with powerful AIs have taken over control of many of its most influential and popular destinations, notwithstanding that they achieve this by enabling (almost) anyone to share their views, creations, and reactions with (almost) anyone.

### 7.8 Non-maleficence

Non-maleficence, and its companion of beneficence, are the two most common and fundamental principles in any ethical decision, almost synonymous with, respectively, "don't be evil" and "be good". Downes (2026) rightly points out that there are occasions when we deliberately intend to cause harm, but that is exactly why we need ethical principles of this nature. We hope that those who choose legal and destructive acts (e.g. wage wars, force people into environmentally destructive commutes, or incarcerate them) have weighed up the principle of non-maleficence with whatever reasons they intend to cause harm. Outside of psychopathic behaviour or crimes of passion, most intentional harm tends to be justified by the aggressor, for instance to protect the rest of the population, to discourage similar behaviour, to protect property, and so on. Again, "all else being equal" is the key to understanding departures from these principles.

### 7.9 Sustainability

Downes (2026) makes some good points in this section, ending with some wittily expressed but truly wicked challenges that, for us, are what make principles of this nature worthwhile, because different principles are explicitly pitted against one another, showing the contours of the arguments in sharp relief. However, the arguments towards the conclusion are notable by their absence. If anything, it appears that Downes does accept this as an ethical principle (we agree, it is), his main argument against it being that many other things we do are worse. If anything this principle is likely to become more important and more strongly connected to non-maleficence as the impacts of non-sustainability impact human experiences.

## 8 Further issues

To educate is to bring about change in another person, for better or for worse, so the issues are inextricably moral. Equally, education embeds structural power relationships deep in its core that, for children, are often enforced through legislation and, for older learners, are maintained through the reward structures of credentials. This makes teachers and educational establishments responsible for what transpires as a result, including breaches of trust, cheating, and academic misconduct. In the past, teachers were often able to exercise tight control over the tools, resources, and methods used for learning. This control has been steadily eroding for some decades, especially since the advent of the World Wide Web. Increased use of learning management systems and other virtual learning environments has further loosened the control of the individual teacher, shunting many ethical and pedagogical decisions to other places in the organizational hierarchy and beyond. Analytics tools are one face of this, as are other AI manifestations such as adaptive hypermedia or intelligent tutoring systems. In many cases, such tools are instruments of power, wielded with beneficent intent but frequently causing harms such as loss of intrinsic motivation, feelings of inadequacy, sub-optimal learning, and lost love of the subject (Dron, 2007). Equally, we recognize many positive ways that AIs can empower learners to take greater control of their own learning (Achuthan, 2025), or enhance social learning (Sethi & Jain, 2024; Zhou & Schofield, 2024).

The issue is complicated by the fact that, more recently, popular forms of GenAI are often not parts of the same power structure, whether tasks are set by teachers or they are independently accessed by students. For teachers who thrive on being in control, we recognize this may be perceived as threatening the very basis of their authority, not to mention undermining their plans, their expertise, and their ability to accurately assess student learning. We think this is a good thing, on the whole, for reasons we have written of extensively in the past, e.g. (Dron, 2007, 2023a). It does, however, raise an ethical concern that neither Downes nor we have so far significantly touched on: that harm, as well as good, may be done to the systems and social

structures in which we participate, and that what may benefit an individual may erode the collective social entities they may be members of, potentially leading to greater harms to more people than the benefits to the individual. As GenAIs are increasingly used in society the opportunity, responsibility and acountability for harms is also emergent and examined. Outside of the scope of this paper we add our weight to larger and urgent discussions arising from irresponsible use of GenAIs and agents and the positioning of accountability for GenAI's abilities. Downes (2026) does begin to allude to this when discussing the trade-offs between raising a billion people out of poverty vs the environmental impacts of doing so, but the problem is far more ubiquitous and of great significance at a local level in education. For instance, for any single individual it may make a great deal of sense to make use of GenAIs to fill in the gaps that classroom teaching leaves unfilled, and the majority of students are now doing so (Freeman, 2025; Ravšelj et al., 2025) but, in so doing, they may dilute and even disable the pedagogies that are in use in that classroom, especially when they rely on whole-class engagement. Such problems are usually easy to fix, which is why we see it as a good thing that teacher control is undermined: it forces teachers to pay more attention to what their students are doing and to adapt, with consequent benefits to motivation, greater learning, and a greater sense of being valued, improving the chances of persistence and success, however that may be measured.

# 9 Conclusion

In clarifying what is meant by AI and considering the relationship between principles, policy and power, we have been able to reply to and build on Downes's (2026) article. "AI" is too general a term to set a context for Downes's paper; the issues raised are primarily those of GenAIs such as large language models and diffusion models for image generation. We agree that GenAI does exhibit at least the outward appearance of intelligence, and that this has impacts on education and society. We note that in considering an educational setting, principles that refer to the use of GenAI must also align with principles of education which take care to reflect the power dynamics in education, as well as the formative and relational nature of education.

Ethics cannot, however, be neatly pinned down to just a set of principles, nor is it, as Downes (2026, p. 14) puts it, just "based in our own sense of what's right and what's wrong (itself a product of culture and education and upbringing and experience and reflection)". Rather, it is a dynamic, shifting, evolving and contested set of models, tastes, values, and attitudes that is situated in and between the relationships we have with one another, within our various cultures, within our communities, within our societies, within organizational structures, with tools, with the world around us, and with ourselves. We should not expect consistency from one entity to the next, nor stability over time, but it is worthwhile to tame the complexity enough that we can understand and talk about the issues. Principles provide a means to do this, acting as boundary objects to support negotiation of meaning. They are are not all that matters, but they matter.

We disagree with Downes that defining principles is intentionally political but we do agree that expressions of principles may be used coercively as well as to empower ethical dialogue: they are largely tools, not complete technologies in their own right, that provide a necessary common language for framing or understanding of ethical behaviour.

Whatever your preferences may be for ethical frameworks and theories, at the heart of almost all of them are two principles: first, to avoid harm and, second, to do good, whether it is for those we care about, anyone at all, our communities, our environments, our fellow creatures, or ourselves. Beyond those two basic ethical principles may be further principles, acting as rules of thumb, guides to action, justifications for rules, and, above all, means of structuring thought and

conversation. Even if we disagree on what the principles mean or whether we should adhere to them, they are starting points for dialogue and reflection. We think this is good.

## References


Achuthan, K. (2025). Artificial intelligence and learner autonomy: A meta-analysis of self-regulated and self-directed learning. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1738751

Adams, C., Pente, P., Lemermeyer, G., & Rockwell, G. (2023). Ethical principles for artificial intelligence in K-12 education. *Computers and Education: Artificial Intelligence*, *4*, 100131. https://doi.org/10.1016/j.caeai.2023.100131

Aristoteles. (1991). The Nicomachean ethics (W. D. Ross, J. L. Ackrill, & J. O. Urmson, Trans.). Oxford University Press. (Original work published 1925, Oxford University Press)

Barcaui, A. (2025). ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention. *Social Sciences & Humanities Open*, *12*, 102287. https://doi.org/10.1016/j.ssaho.2025.102287

Bingham, C. W., & Sidorkin, A. M. (2004). *No education without relation* (Vol. 259). Peter Lang.

Buber, M. (1947). *Between Man and Man*. Routledge.

Cheng, M., Lee, C., Khadpe, P., Yu, S., Han, D., & Jurafsky, D. (2026). Sycophantic AI decreases prosocial intentions and promotes dependence. *Science*, *391*(6792), eaec8352. https://doi.org/10.1126/science.aec8352

China Association for International Science and Technology Cooperation. (2017). *Next Generation Artificial Intelligence Development Plan Issued by State Council*. Department of International Cooperation Ministry of Science and Technology.

Downes, S. (2016, February 17). *Personal and personalized learning*. https://www.downes.ca/post/65065

Downes, S. (2022). Connectivism. *Asian Journal of Distance Education*, *17*(1). https://www.asianjde.com/ojs/index.php/AsianJDE/article/view/623

Downes, S. (2026). On ethical AI principles. *Journal of Open, Distance, and Digital Education*, *3*(1), 1-18. https://doi.org/10.25619/622a7242

Dron, J. (2007). *Control and constraint in e-Learning: Choosing when to choose*. Idea Group International.

Dron, J. (2023a). *How education works: Teaching, technology, and technique*. AU Press.

Dron, J. (2023b). The human nature of generative AIs and the technological nature of humanity: Implications for education. *Digital*, *3*(4), 319–335. https://doi.org/10.3390/digital3040020

Dworkin, R. M. (1967). The model of rules. *The University of Chicago Law Review*, *35*(1), 14–46.

Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., Maes, P., Phang, J., Lampe, M., Ahmad, L., & Agarwal, S. (2025). *How AI and human behaviors shape psychosocial effects of chatbot use: A longitudinal randomized controlled study* (arXiv:2503.17473). arXiv. https://doi.org/10.48550/arXiv.2503.17473

Fernandes, D., Villa, S., Nicholls, S., Haavisto, O., Buschek, D., Schmidt, A., Kosch, T., Shen, C., & Welsch, R. (2026). AI makes you smarter but none the wiser: The disconnect between performance and metacognition. *Computers in Human Behavior*, *175*, 108779. https://doi.org/10.1016/j.chb.2025.108779

Freeman, J. (2025). *Student Generative AI Survey 2025* (Kortext 61). Higher Education Policy Institute.

Friedman, L. M. (2007). *Guarding life's dark secrets: Legal and social controls over reputation, propriety, and privacy*. Stanford University Press.

*General Data Protection Regulation (GDPR) Compliance Guidelines*. (2018). GDPR.Eu. https://gdpr.eu/

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, *15*(1), 6. https://doi.org/10.3390/soc15010006

Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Harvard University Press.

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, *345*(6202), 1340–1343. https://doi.org/10.1126/science.1251560

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Buckingham Shum, S., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, *32*, 504–526. https://doi.org/10.1007/s40593-021-00239-1

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kalai, A. T., Nachum, O., Vempala, S. S., & Zhang, E. (2025). *Why Language Models Hallucinate* (arXiv:2509.04664). arXiv. https://doi.org/10.48550/arXiv.2509.04664

Kant, I. (2017). Fundamental principles of the metaphysic of morals (T. K. Abbott, Trans.; Vol. 8). Delphi Classics. (Original work published 1785).

Kohn, A. (2015). *Four reasons to worry about "personalized learning."*

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task* (arXiv:2506.08872). arXiv. https://doi.org/10.48550/arXiv.2506.08872

Marx, K. (1994). A contribution to the critique of Hegel's philosophy of right: Introduction (R. A. Davis, Trans.). In J. J. O'Malley (Ed.), Marx: Early political writings (pp. 57–70). Cambridge University Press. (Original work published 1844).

Mehta, U. S. (1999). *Liberalism and empire: A study in nineteenth-century British liberal thought*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/L/bo3623192.html

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*(3), 438–450. https://doi.org/10.1111/bjet.12152

Prinsloo, P., & Slade, S. (2017). Ethics and learning analytics: Charting the (un)Charted. In C. Lang, G. Siemens, A. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics* (pp. 49–57). Society for Learning Analytics Research (SoLAR). https://doi.org/10.18608/hla17.004

Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., Iahad, N. A., Abdulla, A. A., Akopyan, A., Segura, M. W. A., AlHumaid, J., Allam, M. F., Alló, M., Andoh, R. P. K., Andronic, O., Arthur, Y. D., Aydın, F., Badran, A., Balbontín-Alvarado, R., Saad, H. B., … Aristovnik, A. (2025). Higher education students' perceptions of ChatGPT: A global study of early reactions. *PLOS ONE*, *20*(2), e0315011. https://doi.org/10.1371/journal.pone.0315011

Selwyn, N. (2019). *Should robots replace teachers?* Polity Press. https://eric.ed.gov/?id=ED599636

Sethi, S. S., & Jain, K. (2024). AI technologies for social emotional learning: Recent research and future directions. *Journal of Research in Innovative Teaching & Learning*, *17*(2), 213–225. https://doi.org/10.1108/JRIT-03-2024-0073

Shiu, P. K., Sterne, G. R., Spiller, N., Franconville, R., Sandoval, A., Zhou, J., Simha, N., Kang, C. H., Yu, S., Kim, J. S., Dorkenwald, S., Matsliah, A., Schlegel, P., Yu, S., McKellar, C. E., Sterling, A., Costa, M., Eichler, K., Bates, A. S., … Scott, K. (2024). A Drosophila computational brain model reveals sensorimotor processing. *Nature*, *634*(8032), 210–219. https://doi.org/10.1038/s41586-024-07763-9

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, *631*(8022), 755–759. https://doi.org/10.1038/s41586-024-07566-y

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1509–1528. https://doi.org/10.1177/0002764213479366

Star, S. L., & Griesemer:, J. R. (1989). Institutional ecology, "translations" and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, *19*(3), 387–420. http://www.jstor.org/stable/285080

Thomas, L. (2026, March 23). *"Our product is used, on occassion, to kill people": Inside Palantir, the world's scariest AI company*. BBC Science Focus Magazine. https://www.sciencefocus.com/future-technology/inside-palantir-the-worlds-scariest-ai-company

Thomson, J. J. (1984). The trolley problem. *Yale LJ*, *94*, 1395.

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence | UNESCO*. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

Villani, C., Bonnet, Y., Schoenauer, M., Berthet, C., Levin, F., Cornut, A. C., & Rondepierre, B. (2018). *For a meaningful artificial intelligence: Towards a French and European strategy*. Conseil national du numérique. https://books.google.ca/ books?id=9cVUDwAAQBAJ

Warr, M., & Heath, M. K. (2025). Uncovering the hidden curriculum in generative AI: A reflective technology audit for teacher educators. *Journal of Teacher Education*, *76*(3), 245–261. https://doi.org/10.1177/00224871251325073

Zhou, X., & Schofield, L. (2024). Using social learning theories to explore the role of generative Artificial Intelligence (AI) in collaborative learning. *Journal of Learning Development in Higher Education*, (30). https://doi.org/10.47408/jldhe.vi30.1031

## Author's Contributions (CRediT)

JD & SG: Conceptualization, Formal analysis, investigation, methodology, project administration, writing – original draft, Writing – review & editing. All authors have read and agreed to the published version of the manuscript.

## Competing Interests

The author has no competing interests to declare.

## Acknowledgement of Use of Generative AI Tools

GenAI was not used in the development of this article.