ASCILITE 2026

The Future is Connection:

Collaboration, Creativity and Compassion in a Digital Society

Supporting ethical engagement with AI in education: Evaluating the ethical coverage of the Scaffolded AI Literacy (SAIL) framework

David Parsons
academyEX

Kathryn MacCallum
University of Canterbury

Ethical engagement with AI in education requires more than awareness of risks or compliance with institutional policy. Educators and learners need to recognise, evaluate, mitigate, and respond to ethical issues as AI tools become embedded in learning and teaching. This paper examines whether the Scaffolded AI Literacy (SAIL) framework provides a sufficiently comprehensive set of competencies for this purpose. Six cross-sector systematic reviews and mappings of AI ethics are analysed to provide a benchmark of recurring ethical issues identified in relation to AI. The comparison shows strong coverage in the SAIL framework of issues including fairness, equity, transparency, privacy and safety. Human agency, accountability, well-being, human rights, and participation are partially addressed but less explicit. Notably, environmental sustainability demonstrates the clearest gap between emphasis in the literature and explicit coverage in the framework. The analysis supports SAIL as a substantial framework for ethical AI literacy while identifying targeted opportunities for refinement.

Keywords: AI ethics, AI literacy, SAIL, framework evaluation, sustainability.

Introduction

Ethical concerns about the use of AI in education are a significant challenge for educators and learners. Overarching concerns relate to the overall trustworthiness of AI and the role it may play in potentially amplifying social biases and causing harm (Porayska-Pomsta et al., 2023). Specific issues of concern are many and varied, including governance, transparency, sustainability, privacy, safety, inclusiveness, and human-centredness (Nguyen et al., 2023). Educators need a structured approach to the ethics of AI if they are to be confident in their own ethical approach to teaching and learning with AI tools and assist their students to use AI ethically.

Ethical AI engagement in education is not just the concern of institutional policy or individual caution but can be developed as a scaffolded literacy. The Scaffolded AI Literacy (SAIL) framework (MacCallum et al., 2026), developed by the authors of this paper, offers a way to translate broad ethical AI concerns into staged competencies for learners and educators, building capability through awareness, application, and critical evaluation.

We have been applying the SAIL framework across a range of contexts in an effort to assess its utility as a support for AI literacy development. Part of that process has been to refine the framework by identifying areas for improvement. The framework was developed from a Delphi study in 2023, and the evolution of AI since then raises ongoing questions about its currency. Significant updates have been made to the framework (MacCallum et al., 2026), but as part of a process of continuing review, this paper examines a literature synthesis in the area of AI ethics as a benchmark against which relevant parts of the framework can be examined. The question for the study was: To what extent does the framework provide coverage of ethical concerns that recur across syntheses of AI ethics?

Establishing an AI ethics benchmark

According to Kazim and Koshiyama (2021), there are three dominant theories of ethics: utilitarianism, rights, and virtue. Each of these ethical theories implies distinct AI literacy needs in education. From the utilitarian

perspective, ethics is about weighing the consequences of actions. Rights focus on the entitlements and protections owed to individuals and groups, while the domain of virtue relates to values. Ethical AI literacy therefore needs to extend beyond awareness of rules toward reasoning, judgement and action.

To gain an insight into the key ethical issues identified in the context of AI, a set of literature syntheses was analysed to identify the most common issues. The inclusion criteria were journal-published systematic reviews or mappings where AI ethics was central to the study, there was a cross-domain scope, and the findings provided an identifiable synthesis of ethical issues. This was a targeted synthesis rather than a systematic review of reviews. Initially, six systematic reviews or mappings were selected. In each case, AI ethics was the focus of their analysis, and their findings were intended to apply across domains. Each of the selected articles has a different focus. Jobin et al. (2019) mapped AI ethics guidelines to see whether a global agreement on ethical AI is emerging. Ashok et al. (2022) developed an AI ontology with governance at its centre and applied it to a set of digital technology archetypes. Laine et al. (2024) explored the knowledge contributions (cognitive, pragmatic, and normative) that auditing research provides to various stakeholder groups. Ortega-Bolaños et al. (2024) reviewed a set of conceptual and technical tools and applied them to specific stages of the AI lifecycle. Palumbo et al. (2025) looked for practical, objective metrics that can be integrated into data pipelines. Bach et al. (2025) focused on empirical studies of AI applications and areas of responsible AI practices. This diverse selection of articles provided a benchmark against which to assess the ethical coverage of the SAIL framework.

Assessing coverage of ethical issues in the SAIL framework

The SAIL framework provides competencies across three domains: AI Concepts, Application of AI and Technical Skills, and AI Digital Citizenship (MacCallum et al., 2026). The latter domain includes two categories, Social, Cultural, and Ethical Issues and Risks and Mitigations. Across these two categories, 10 key ideas are scaffolded across four levels, each focusing on increasing depth of literacy. While most of the ethical issues identified in the selected literature were linked with the AI Digital Citizenship domain, relevant key ideas also occur elsewhere in the framework, such as AI Transparency in Application of AI and Technical Skills, so the analysis considers all three domains.

Table 1 provides a summary of 10 principles drawn from the six articles addressing ethical concerns associated with AI. Rather than treating ethical concerns as a catalogue of risks or negative consequences associated with AI, we conceptualise them as AI ethics principles: high-level normative values that provide a basis for considering how AI should be understood, used and governed. Such principles encompass both the mitigation of potential harms and the promotion of desirable outcomes, shifting attention from what might go wrong with AI towards broader questions of what ethical and socially beneficial engagement with AI should entail (Jobin et al., 2019). This framing is particularly relevant to AI literacy, as it enables learners to recognise ethical issues while also developing the capacity to reason about competing values, possible benefits and appropriate courses of action. In column two of Table 1, each principle is mapped to relevant key ideas of the SAIL framework. In the final column, coverage is classified as strong where the principle is explicit and progressively developed, partial where relevant competencies exist but important aspects of coverage or progression are implicit, or absent. Potential refinements are included in this column.

It is important to note that in constructing this table, it has been necessary to make a series of judgments about the equivalence of the naming of ethical issues. Column one of the table contains commonly identified terms from the sources, but terminology varies. In the analysis, issues were coded as broadly equivalent where there were clearly matching themes in the synthesised findings of the sources. For example, the issue of beneficence, well-being, and social benefit subsumes related terms such as social/common good, human flourishing, and value creation.

Table 1.
Coverage of literature-derived ethical concerns in SAIL

Ethical principles	Related key ideas from SAIL	Assessment/potential refinement
Fairness and equity[1,2,3,4,5,6]	AI and Equity, AI Bias, and AI and Cultural Diversity address unequal impacts, accessibility, bias, structural inequities and inclusive design.	Strong. Could broaden higher-level competencies to include intersectional bias and how different definitions of fairness may produce different outcomes.
Transparency and explainability[1,2,3,4,5,6]	AI Transparency addresses explainability, accountability and ethical governance.	Strong. No major change required.
Privacy and data governance[1,2,3,4,5,6]	Risks of AI Systems addresses privacy and misuse. AI Policy addresses governance, oversight, and regulation, including obligations to Indigenous rights and contributing to policy development at the higher levels.	Strong. Could more explicitly address how personal data may be misused, for example to reveal sensitive information or identify individuals.
Safety and non-maleficence[1,2,3,4,5,6]	Safe and Responsible Use of AI recommends safety-first practices, Mitigating the Risks of AI addresses reliability and Risks of AI Systems addresses security.	Strong. Could make monitoring after implementation, uncertainty, and unexpected harms more explicit.
Human agency and autonomy[1,2,3,4,5,6]	Mitigating the Risks of AI addresses consent, and Critical Application of AI and Safe and Responsible Use both encourage critical engagement and human decision-making.	Partial. Could include how AI affects human agency and when human intervention, refusal or override is required.
Accountability and responsibility[1,2,3,4,5,6]	AI Policy addresses governance and oversight; AI Transparency addresses accountability; Safe and Responsible Use assigns responsibility for impacts.	Partial. Could add the ability to examine, challenge and correct AI-supported decisions and provide routes for redress.
Beneficence, well-being, and social benefit[1,2,3,4,5,6]	AI and Society considers benefits and challenges, while Critical Application of AI considers whether AI is appropriate for a task. Equitable outcomes are addressed in Goal-based AI Usage and Safe and Responsible Use of AI.	Partial. Could address the need to evaluate whether AI use produces worthwhile benefits, who benefits, and who may be disadvantaged.
Inclusion and participation[1,2,4]	Participation is addressed in both AI and Society and Engagement with AI. Inclusion is implicit in AI and Equity and AI and Cultural Diversity.	Partial. Coverage could be revised to more explicitly address the concept of inclusion.
Human rights and human dignity [1,4,5,6]	The framework touches on personal, Indigenous, and minority rights in AI Policy and Mitigating the Risks of AI.	Partial. Human dignity and broader rights issues could be better addressed, noting the primacy of rights in ethics.
Environmental sustainability[1,2,5]	Safe and Responsible Use refers to long-term consequences but no explicit mention of sustainability.	Absent. Environmental impact is not explicitly mentioned in the framework.

Sources: Jobin et al. (2019)[1], Ashok et al. (2022)[2], Laine et al. (2024)[3], Ortega-Bolaños et al. (2024)[4], Palumbo et al. (2025)[5], and Bach et al. (2025)[6].

Findings

The SAIL framework provides substantial coverage of the concerns most consistently identified in the cross-sector literature represented by the selected syntheses. Fairness, transparency, privacy, and safety are strongly represented. This supports the suitability of the framework for helping learners to develop ethical capability. However, there are also areas where the competencies should be further developed.

One significant gap identified through the mapping was environmental sustainability. Although SAIL addresses a broad range of social, cultural, ethical and risk-related dimensions of AI, environmental sustainability is not currently represented as a distinct key idea. This aspect of AI ethics is increasingly important as the environmental implications of AI become more visible. AI systems have environmental impacts across their lifecycle, including energy and water consumption, carbon emissions, material and infrastructure demands, and electronic waste (Schwartz et al., 2020). At the same time, AI can also contribute positively to environmental sustainability, including through climate modelling, energy optimisation and other applications supporting sustainable development (Kaack et al., 2022; Vinuesa et al., 2020). Sustainability should therefore not be framed simply as another "risk" of AI, but as an ethical principle requiring learners to consider both the environmental costs and potential environmental benefits of AI.

This issue is increasingly relevant in education as some students question or reject AI use on ethical grounds, including concerns about environmental sustainability. Such "conscientious objection" highlights that responsible AI literacy should include the capacity to make informed decisions not only about how to use AI, but whether to use it at all. This is complicated by the opacity of AI's environmental footprint, with estimates varying by model, task, infrastructure, and energy source, and often relying on incomplete information (Crawford, 2024). Rather than relying on fixed estimates, learners therefore need to be able to critically interpret sustainability claims, recognise uncertainty, and make proportionate decisions about AI use.

Based on these concerns and the lack of current coverage in the framework, we propose adding the following competency progression to the SAIL framework as key idea 30: AI and Environmental Sustainability.

SAIL level 1: Understand and Explore AI
Explain that AI systems have environmental impacts across their lifecycle, including energy and water consumption, carbon emissions, resource use and electronic waste.

SAIL level 2: Apply and Integrate AI
Make informed and proportionate choices about when and how to use AI, considering its environmental impacts and applying practical strategies to reduce unnecessary resource use while still achieving the intended purpose.

SAIL level 3: Evaluate and Create AI
Critically evaluate evidence and claims about the environmental impacts and benefits of AI, considering lifecycle effects, uncertainty, transparency and trade-offs, and design or adapt AI-enabled practices to support sustainable outcomes.

SAIL level AI++: Transform and Develop AI (Beyond AI Literacy)
Develop, optimise or influence AI systems, practices, policies or infrastructure to advance environmental sustainability, considering wider systemic consequences and trade-offs.

Conclusion

Ethical engagement with AI in education can be understood as a developmental AI literacy challenge, but a framework intended to support it should be tested against the ethical concerns identified in the wider literature. This article has provided an exploratory coverage evaluation of the SAIL framework, using recurring concerns identified in cross-sector AI ethics syntheses as an external benchmark, which both supports areas of the framework and identifies areas requiring refinement. Currently, the SAIL framework provides strong coverage of many recurrent concerns, but our analysis has revealed others that are not adequately addressed. Returning to the discussion of ethical theories, some aspects of rights and values could be strengthened by making several areas of human agency, accountability, beneficence, inclusion and human rights more explicit. However, environmental sustainability is the clearest gap in the current explicit coverage of the framework. The analysis therefore supports SAIL as a substantial foundation for ethical AI literacy while identifying focused opportunities for refinement.

References

Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for artificial intelligence and digital technologies. *International Journal of Information Management, 62*, 102433. https://doi.org/10.1016/j.ijinfomgt.2021.102433

Bach, T. A., Kaarstad, M., Solberg, E., & Babic, A. (2025). Insights into suggested responsible AI (RAI) practices in real-world settings: A systematic literature review. *AI and Ethics, 5*, 3185–3232. https://doi.org/10.1007/s43681-024-00648-7

Crawford, K. (2024). Generative AI's environmental costs are soaring—and mostly secret. *Nature, 626*, 693. https://doi.org/10.1038/d41586-024-00478-x

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

Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change, 12*, 518–527. https://doi.org/10.1038/s41558-022-01377-7

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns, 2*(9), 100314. https://doi.org/10.1016/j.patter.2021.100314

Laine, J., Minkkinen, M., & Mäntymäki, M. (2024). Ethics-based AI auditing: A systematic literature review on conceptualizations of ethical principles and knowledge contributions to stakeholders. *Information & Management, 61*(5), 103969. https://doi.org/10.1016/j.im.2024.103969

MacCallum, K., Parsons, D., & Mohaghegh, M. (2026). The Scaffolded AI literacy (SAIL) framework: Results of a Delphi study for equitable AI literacy framework design in education. *Computers and Education: Artificial Intelligence, 10*, 100584. https://doi.org/10.1016/j.caeai.2026.100584

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies, 28*(4), 4221-4241. https://doi.org/10.1007/s10639-022-11316-w

Ortega-Bolaños, R., Bernal-Salcedo, J., Germán Ortiz, M., Galeano Sarmiento, J., Ruz, G. A., & Tabares-Soto, R. (2024). Applying the ethics of AI: A systematic review of tools for developing and assessing AI-based systems. *Artificial Intelligence Review, 57*, 110. https://doi.org/10.1007/s10462-024-10740-3

Palumbo, G., Carneiro, D., & Alves, V. (2025). Objective metrics for ethical AI: A systematic literature review. *International Journal of Data Science and Analytics, 20,* 247–267. https://doi.org/10.1007/s41060-024-00541-w

Porayska-Pomsta, K., Holmes, W., & Nemorin, S. (2023). The ethics of AI in education. In B. Du Boulay, A. Mitrovic, & K. Yacef (Eds.), *Handbook of artificial intelligence in education* (pp. 571-604). Edward Elgar Publishing. https://doi.org/10.4337/9781800375413.00038

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM, 63*(12), 54–63. https://doi.org/10.1145/3381831

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications, 11*, 233. https://doi.org/10.1038/s41467-019-14108-y

Parsons, D., & MacCallum, K. (2026). Supporting ethical engagement with AI in education: Evaluating the ethical coverage of the Scaffolded AI Literacy (SAIL) framework. In Cochrane, T., Narayan, V., Poole, T., Lai, P., Lake, P., Moore, O., & Li, L. (Eds.), *The Future is Connection: Collaboration, Creativity and Compassion in a Digital Society*. Proceedings ASCILITE 2026. Brisbane (pp. 780–784). https://doi.org/10.65106/apubs.2026.3164

Note: All published papers are refereed, having undergone a double-blind peer-review process.
The author(s) assign a Creative Commons by attribution license enabling others to distribute, remix, tweak, and build upon their work, even commercially, as long as credit is given to the author(s) for the original creation.

© Parsons, D. & MacCallum, K. 2026