Beyond Engagement: A Human Flourishing Framework for Emotion-Aware AI in Education

Spyridon Kontis and Sofia Anastasiadou
Department of Midwifery, University of Western Macedonia, Kozani, Greece

dmw00034@uowm.gr
sanastasiadou@uowm.gr

Abstract: Emotion-aware artificial intelligence (AI) is increasingly used to infer learners' affective states, such as frustration, confusion, or engagement, and to adapt instruction in response. These capabilities may make learning environments more responsive, but they also raise questions about privacy, surveillance, bias, and the treatment of emotional experience as data. Evaluation has therefore moved beyond accuracy alone. Trustworthy AI stresses transparency and oversight; AI Ethics adds fairness and harm prevention; Human-Centred AI foregrounds learner agency. These perspectives are necessary, yet they do not fully settle a further educational question: against what longer-term human purpose should emotion-aware AI be judged? This conceptual paper proposes the Human Flourishing Framework (HFF) as a normative evaluative perspective for addressing that question. It is not presented as a computational layer or a causal model. Instead, it examines the conditions created around AI-mediated learning within a wider educational ecosystem. The framework organises five dimensions, Emotional Autonomy, Psychological Safety, Transparency and Data Dignity, Fairness and Inclusion, and Human Flourishing, across Foundation, Empowerment, and Outcome layers. Its purpose is to examine whether the design and use of emotion-aware AI support or constrain learners' longer-term development, without attributing flourishing to the technology itself. The paper clarifies the framework's epistemological and theoretical positioning, distinguishes flourishing from related concepts such as capability, agency, and self-regulation, and translates the HFF dimensions into indicative system requirements. The contribution is conceptual rather than empirical. The framework has not yet been validated or operationalised as a measurement model; future work will therefore require longitudinal, multi-level, and cross-cultural research capable of studying AI-mediated learning within its wider educational context.

Keywords: Emotion-Aware Artificial Intelligence; Human Flourishing Framework; Trustworthy AI; AI Ethics; Human-Centred AI; Educational Technology Evaluation

1. Introduction

Artificial intelligence increasingly attempts to interpret how learners feel, not only what they know. Affect-aware systems use facial, vocal, textual, physiological, and interaction cues to infer states that may matter for learning, drawing on a long tradition of affective computing and research on achievement emotions (Picard 1997; Calvo and D'Mello 2010; Pekrun 2006).

In educational settings, such inferences can inform the timing of feedback, pacing, or support during frustration (Grawemeyer et al. 2017). The same capability is also difficult to treat as neutral. Emotional expression is context-sensitive and can be interpreted incorrectly, while the collection and inference of affective data may create surveillance and datafication concerns (Barrett et al. 2019; McStay 2020; Williamson and Eynon 2020).

As these concerns have become more visible, evaluation has widened from technical performance to questions of trust, ethics, and human agency. Trustworthy AI emphasises reliability, transparency, privacy, and oversight; AI Ethics adds questions of fairness, dignity, and justified use; Human-Centred AI asks whether people remain meaningful participants in AI-mediated decisions (European Commission High-Level Expert Group on Artificial Intelligence 2019; UNESCO 2021; Shneiderman 2020). These perspectives overlap rather than forming a simple sequence, and each addresses an important part of responsible educational AI.

A further question nevertheless remains. A system may be accurate, transparent, and ethically governed while leaving uncertain whether its sustained use is educationally worthwhile for the learner over time. This is especially important for emotion-aware AI, because repeated interpretation of affect can shape how learners understand their own responses, when they seek help, and how much authority they give to automated feedback (McStay 2020; Kontis and Anastasiadou 2026a).

Against this background, the paper asks: How can human flourishing provide a normative evaluative lens for emotion-aware AI in education beyond technical performance, ethical compliance, and short-term engagement? The Human Flourishing Framework (HFF) is proposed as a conceptual response. It builds on, rather than replaces, existing approaches and shifts attention toward the educational conditions surrounding AI use and their relevance to learners' longer-term development. Figure 1 summarises the conceptual move that motivates the framework.

2. Conceptual Foundations

2.1 Emotion-Aware AI in Education

Emotion-aware AI rests on a simple but consequential premise: learning is not purely cognitive. Curiosity, frustration, anxiety, and other affective states influence attention, persistence, and the way learners engage with difficult material (Pekrun 2006; D'Mello and Graesser 2012). Systems may infer such states from facial, vocal, textual, behavioural, or physiological signals and adapt their responses accordingly (Calvo and D'Mello 2010).

The inference itself, however, is probabilistic and context-dependent. Facial movement, for example, cannot be mapped mechanically onto discrete emotions across individuals and cultures (Barrett et al. 2019). An adaptive response based on a confident but mistaken interpretation may therefore be pedagogically inappropriate even when the underlying model performs well. Educational value depends not only on detection accuracy, but on how affective interpretations are situated, communicated, and used.

2.2 Trustworthy AI

Trustworthy AI provides a necessary baseline for this discussion. Robustness, transparency, privacy, accountability, and human oversight matter whenever automated inferences influence educational decisions (European Commission High-Level Expert Group on Artificial Intelligence 2019; Floridi and Cowls 2019; Anastasiadou 2026a). In emotion-aware learning, these requirements are particularly important because learners may not know that an affective inference has been made or how it has shaped the system's response. Trustworthiness helps establish whether reliance on a system is justified and governed, but it does not by itself specify the broader educational end that such reliance should serve.

2.3 AI Ethics and Human-Centered AI

AI Ethics extends the discussion to fairness, dignity, harm, and the justification of deployment. These issues become sharper when the data concern emotion, because inferred affect can be intimate, uncertain, and culturally variable (McStay 2020; Barrett et al. 2019; UNESCO 2021). Human-Centred AI adds another important concern: whether learners and educators can understand, question, and influence AI-mediated decisions instead of becoming passive recipients of them (Amershi et al. 2019; Shneiderman 2020).

Taken together, these perspectives establish strong protective and participatory conditions. Yet preserving agency within an interaction does not automatically answer what that interaction should contribute to over months or years of learning. The question of educational purpose remains, which is where human flourishing becomes relevant.

2.4 Human Flourishing

Human flourishing is used here as a broader developmental idea, not as a synonym for wellbeing, capability, autonomy, or self-regulation. The Capability Approach asks what people are genuinely able to do and become, rather than judging development only by resources or outcomes (Sen 1999; Nussbaum 2011). Self-Determination Theory identifies autonomy, competence, and relatedness as psychological conditions that support motivated development (Ryan and Deci 2020). Agency refers here to a learner's capacity to act, choose, question, and influence an educational interaction, whereas self-regulation concerns the more specific management of cognition, behaviour, and emotion. These concepts overlap, but they operate at different levels.

Flourishing provides the wider educational horizon within which these capacities acquire significance. Positive Education and recent policy work on education for human flourishing place meaning, development, relationships, and participation alongside academic achievement (Seligman et al. 2009; OECD 2025). In this paper, flourishing refers to the sustained development of learners as capable, autonomous, reflective, and purposeful participants in education and society. A system might improve short-term engagement or help regulate frustration without contributing to that broader trajectory. This distinction is central to the HFF: flourishing is an evaluative horizon, not a score or state produced by an AI system.

2.5 Current Educational AI Evaluation

Educational AI is commonly evaluated through measures such as performance, engagement, usability, trust, and fairness, all of which remain useful (Vistorte et al. 2024). For emotion-aware AI, however, short-term indicators

can leave important questions open. Does support become dependence? Can an affective label be challenged? Do repeated adaptations preserve the learner's own judgement? A flourishing-oriented perspective adds these questions of purpose and time without displacing established technical and ethical criteria.

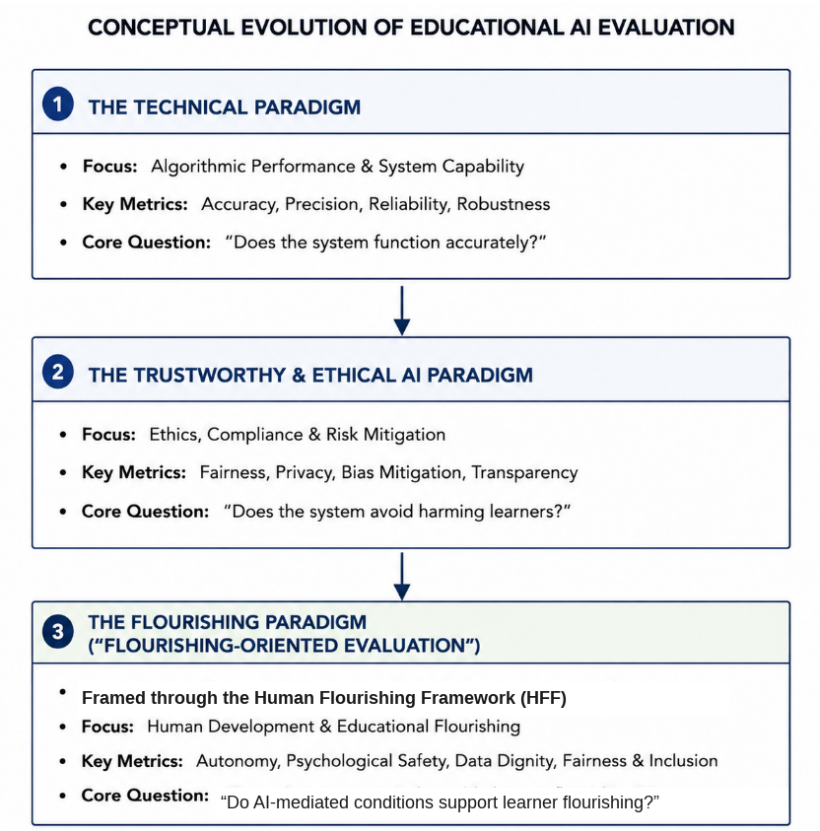


Figure 1: Conceptual evolution of educational AI evaluation from technical performance toward the Human Flourishing Framework

3. The Human Flourishing Framework

3.1 Why a Human Flourishing Framework?

The rationale for the HFF is not that existing frameworks have failed. They answer different questions. Trustworthy and ethical approaches ask whether systems are reliable, governed, and acceptable; human-centred approaches ask whether people retain meaningful agency. Human flourishing adds a longer educational horizon, asking whether the conditions associated with AI use are compatible with learners becoming more capable, autonomous, and able to pursue meaningful learning over time (Nussbaum 2011; Ryan and Deci 2020; OECD 2025).

That question cannot be located in an isolated learner-AI relationship. Educational development takes place within a wider socio-technical ecosystem involving learners, educators, pedagogical choices, institutional arrangements, governance, and cultural context. Emotion-aware AI is one element of that ecosystem, not an independent developmental agent. The HFF therefore appraises AI-mediated educational conditions within this broader setting and does not assume that learner flourishing can be causally attributed to the technology alone.

3.2 Conceptual Architecture of the Human Flourishing Framework

The HFF is not intended as a checklist or causal pathway. It is a layered architecture for organising questions about the conditions under which emotion-aware AI is designed, used, and evaluated. Flourishing is treated as a developmental orientation that may be supported or constrained within AI-mediated educational contexts, rather than as a property that a system possesses.

Three layers structure the framework. The Foundation Layer brings together the trustworthy and ethical conditions required for responsible use, including transparency, privacy, accountability, oversight, and

governance (European Commission High-Level Expert Group on Artificial Intelligence 2019; UNESCO 2021). The Empowerment Layer examines the learner's lived interaction through four HFF dimensions: Emotional Autonomy, Psychological Safety, Transparency and Data Dignity, and Fairness and Inclusion. The Outcome Layer places Human Flourishing as the longer-term educational horizon against which those conditions are considered.

The layers are related but should not be read as a deterministic sequence. A trustworthy foundation does not guarantee empowerment, and empowerment does not guarantee flourishing. Their value lies in making different evaluative questions visible and in showing where technical, ethical, pedagogical, and developmental concerns meet.

Figure 2 presents this architecture and locates the HFF dimensions within the three layers through which conditions relevant to flourishing are examined.

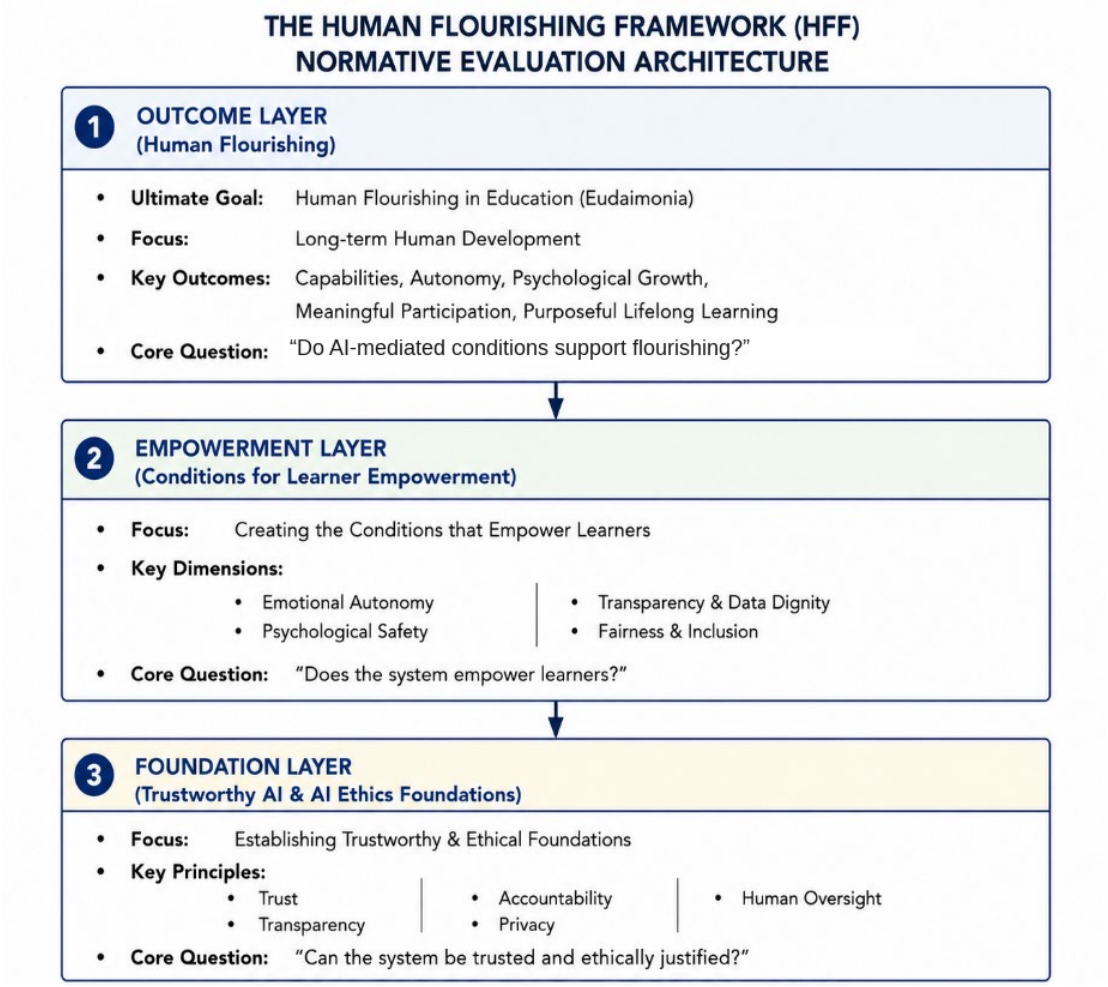


Figure 2: Conceptual architecture of the Human Flourishing Framework (HFF), showing how foundational principles support learner empowerment and flourishing-oreinted evaluation.

3.3 HFF Dimensions

3.3.1 Emotional Autonomy

Emotional Autonomy concerns the learner's capacity to retain meaningful authority over the interpretation and regulation of their own emotional experience during AI-mediated learning. It is distinct from self-regulation, which concerns the management of emotion, and from agency, which refers more broadly to the capacity to act, question, choose, and influence the educational interaction. Emotion-aware AI may support reflection, but an inference should not become a verdict about how a learner feels. Learners should be able to question, override, or disregard affective recommendations when they do not fit their experience (Ryan and Deci 2020; Shneiderman 2020).

3.3.2 Psychological Safety

Psychological Safety asks whether learners can experience uncertainty, difficulty, and error without feeling judged or penalised by affective monitoring. The concept originated in research on interpersonal learning climates, but its underlying concern is relevant when educational systems continuously observe behaviour and infer internal states (Edmondson 1999; McStay 2020). In this setting, safe design means avoiding punitive or high-stakes uses of inferred emotion, limiting unnecessary monitoring, and preserving space for learners to make mistakes without adjusting their behaviour simply to satisfy perceived algorithmic expectations.

3.3.3 Transparency and Data Dignity

Transparency and Data Dignity concern what learners know and can influence about the collection and use of emotional information. Here, data dignity means treating affective data and inferences as closely connected to personhood and vulnerability, rather than as ordinary behavioural features (McStay 2020; UNESCO 2021). Learners should know what is collected or inferred, why it is used, how long it is retained, and how it influences system responses. Transparency is therefore meaningful only when learners can also question or contest an interpretation that affects them.

3.3.4 Fairness and Inclusion

Fairness and Inclusion concern whether emotion-aware AI works equitably across learners whose emotional expression differs by culture, language, context, or neurodiversity. Research on facial expression cautions against assuming that outward behaviour provides a universal reading of internal emotion (Barrett et al. 2019). Fair design should therefore test affective models across diverse populations and provide ways to correct or contest misclassification, rather than treating difference as error.

3.3.5 Human Flourishing

Human Flourishing functions as the framework's evaluative horizon rather than as an additional system output. It concerns whether the wider pattern of AI-mediated learning is compatible with learners' longer-term development as autonomous, capable, reflective, and purposeful participants. The preceding dimensions identify conditions that may support such development, but none is sufficient on its own and no direct causal claim is made. Evidence about flourishing would require attention to time, context, pedagogy, and the wider educational ecosystem (Nussbaum 2011; OECD 2025).

3.4 Epistemological and Theoretical Positioning of the Human Flourishing Framework

The HFF adopts an interpretive and humanistic position toward emotion and educational development. Learners' emotions are not treated as fixed internal states that can be transparently read from data. AI outputs are probabilistic interpretations shaped by training data, modelling choices, behaviour, and context; the learner's own account and the educator's contextual judgement therefore retain epistemic weight (Barrett et al. 2019; McStay 2020). Human flourishing is approached in the same spirit: as a normative educational orientation, not an objective variable generated by an algorithm.

The normative dimension of the HFF is therefore external to the algorithm. It is a perspective used by designers, educators, institutions, and researchers when deciding which objectives should be pursued, which interventions are educationally justified, and which consequences should count as desirable. Normativity resides in these human criteria for design, deployment, interpretation, and evaluation, not in the data or algorithm as such. In this sense, the HFF complements Trustworthy AI, AI Ethics, and Human-Centred AI rather than adding a computational layer to the system (Floridi and Cowls 2019; Shneiderman 2020).

4. Implications

4.1 Implications for Educational AI Design

The practical value of the HFF depends on whether its normative commitments can inform concrete design decisions. It does not provide engineering specifications or universal rules. Instead, each dimension can be translated into an indicative system requirement that helps designers and institutions ask what follows from an affective inference, who can challenge it, and what kind of learner-system relationship repeated use may create. Table 1 summarises these requirements.

Table 1: Indicative system requirements derived from the HFF.

HFF Dimension	Indicative System Requirement
Emotional Autonomy	Allow learners to question, override, or ignore affective interpretations; use AI to support reflection rather than replace learner judgement.
Psychological Safety	Avoid punitive or high-stakes action based on inferred emotion; minimize unnecessary monitoring and conditions that encourage self-censorship.
Transparency and Data Dignity	Explain what affective data are collected or inferred, why they are used, how long they are retained, and how they influence system responses; provide meaningful controls.
Fairness and Inclusion	Test affective interpretation across culturally, linguistically, and neurodiverse populations and provide mechanisms to correct or contest misclassification.
Human Flourishing	Evaluate success beyond engagement or immediate performance, including autonomy, capability development, meaningful participation, and dependence on AI support over time.

For designers, the key question is therefore not only whether an affective inference is accurate, but what educational action follows from it, whether the learner can contest it, and whether repeated use strengthens rather than displaces human judgement (Sadek et al. 2024).

4.2 Implications for Educational Practice

For educational practice, emotion-aware AI is better treated as a source of provisional pedagogical insight than as a substitute for judgement. An affective inference may prompt an educator to look more closely, but it cannot provide the relational and contextual knowledge needed to understand a learner's experience on its own (Holstein and Aleven 2022; Celik 2023). The practical question is therefore not simply whether the system detects frustration, but how a teacher or learner responds to that suggestion and whether support gradually strengthens independent judgement rather than dependence on automated interpretation.

4.3 Implications for Research and Educational AI Governance

For research, the HFF offers a vocabulary for examining developmental questions alongside established measures of performance, engagement, trust, and usability. Its dimensions are not yet validated constructs, and future studies will need to examine them within the wider educational setting rather than attempting to isolate AI as a single cause. Governance can use the framework in a similar way. Existing requirements for safety, privacy, transparency, and accountability remain essential; the HFF adds an external educational-purpose test concerned with whether deployment choices are compatible with learner agency and longer-term development (UNESCO 2021; Miao and Holmes 2023).

5. Contribution, Limitations and Future Research

5.1 Theoretical Contribution

The contribution of this paper is integrative rather than the invention of a new set of ethical principles. The HFF brings Trustworthy AI, AI Ethics, Human-Centred AI, and Human Flourishing into a single educational evaluative architecture and makes two issues more explicit: educational purpose and the time horizon over which AI use is judged. It reframes evaluation from system performance and ethical acceptability alone toward the conditions that AI-mediated learning may create for human development. As a conceptual framework, its claim is one of coherence and usefulness, not empirical superiority over existing approaches.

5.2 Limitations

Several limitations follow from that position. The HFF has not been empirically validated, and its dimensions remain conceptual rather than operationalised measurement constructs. Flourishing is also culturally and philosophically contested; the framework draws substantially on the Capability Approach, Self-Determination Theory, Positive Education, and OECD work, traditions that may not transfer unchanged across educational settings. Nor does the framework imply that flourishing can be causally attributed to an AI system in isolation.

Learner development is shaped by interacting pedagogical, relational, institutional, and cultural factors. Finally, concepts such as Emotional Autonomy and Data Dignity require further operational clarification before they can support comparative measurement.

5.3 Future Research

The immediate research task is to operationalise the HFF dimensions using established measures where appropriate and new items where emotion-aware AI introduces distinct concerns. Longitudinal and multi-level designs are especially important because developmental change unfolds over time and across learner, classroom, and institutional levels. Mixed-method work could combine behavioural or survey measures with learner accounts of how affective interpretations are experienced and contested. Cross-cultural studies are also needed, together with design-based research examining whether the indicative system requirements proposed here remain useful across intelligent tutoring systems, adaptive platforms, and generative AI environments.

6. Conclusion

Emotion-aware AI makes a difficult educational question visible: a system can respond to affect without knowing what that response should ultimately be for. Trustworthy AI, AI Ethics, and Human-Centred AI provide indispensable conditions for responsible use, but educational purpose still needs to be articulated. The Human Flourishing Framework offers one way of doing so while avoiding the claim that technology itself produces flourishing.

Its central shift is from asking only what an intelligent system can do to asking what kinds of educational conditions its design and use help create. Emotional Autonomy, Psychological Safety, Transparency and Data Dignity, and Fairness and Inclusion make that question more concrete, while Human Flourishing keeps the longer developmental horizon in view.

The framework should now be tested, challenged, and refined. Its value will depend less on whether it predicts a single outcome than on whether it helps researchers, educators, designers, and institutions recognise when technically successful systems support, or quietly undermine, the human purposes education is meant to serve.

Ethics Declaration

This conceptual study did not involve human participants, personal data, or animal subjects. Therefore, ethical approval was not required.

AI Declaration

Generative AI tools were used for language editing, structural review, and formatting support. The authors retained responsibility for source selection and verification, conceptual decisions, interpretation of the literature, and the final manuscript.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fournier-Tombs, E., Horvitz, E., Johnston, S., Kikin-Gil, R., McDermott, K., Penrose, C., Morris, M.R. and Olsen, J. (2019) "Guidelines for human-AI interaction", Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, 4-9 May 2019, ACM, New York, pp. 1-13. https://doi.org/10.1145/3290605.3300229

Anastasiadou, S.D. (2026a) "H-IoT Ethics Instrument (H-IoTEI): Development, Validity, and Reliability", in Anastasiadou, S., Seremeti, L. and Papalexandris, S. (eds.), Bioethics and Legal Aspects of AI in Healthcare and Life Sciences, IGI Global Scientific Publishing, Hershey, PA, pp. 225-236. https://doi.org/10.4018/979-8-3373-5646-4.ch010

Barrett, L.F., Adolphs, R., Martinez, A., Ide, D.L. and Russell, J.A. (2019) "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements", Psychological Science in the Public Interest, Vol. 20, No. 1, pp. 1-68. https://doi.org/10.1177/1529100619832930

Calvo, R.A. and D'Mello, S. (2010) "Affect detection and an overview of expressing affective states", IEEE Transactions on Affective Computing, Vol. 1, No. 1, pp. 18-37. https://doi.org/10.1109/TAFFC.2010.2

Celik, I. (2023) "Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education", Computers in Human Behavior, Vol. 138, Article 107468. https://doi.org/10.1016/j.chb.2022.107468

D'Mello, S. and Graesser, A. (2012) "Dynamics of affective states during complex learning", Learning and Instruction, Vol. 22, No. 2, pp. 145-157. https://doi.org/10.1016/j.learninstruc.2011.10.001

Edmondson, A. (1999) "Psychological safety and learning behavior in work teams", Administrative Science Quarterly, Vol. 44, No. 2, pp. 350-383. https://doi.org/10.2307/2666999

European Commission High-Level Expert Group on Artificial Intelligence (2019) Ethics Guidelines for Trustworthy AI, European Commission, Brussels.

Floridi, L. and Cowls, J. (2019) "A unified framework of five core principles for AI in society", Harvard Data Science Review, Vol. 1, No. 1. https://doi.org/10.1162/99608f92.8cd550ed

Grawemeyer, B., Mavrikis, M., Holmes, W., Hansen, A., Loibl, K. and Gutierrez-Santos, S. (2017) "Affective transitions in open-ended learning environments", IEEE Transactions on Learning Technologies, Vol. 10, No. 3, pp. 315-326. https://doi.org/10.1109/TLT.2016.2614810

Holstein, K. and Aleven, V. (2022) "Designing for human-AI complementarity in K-12 classrooms", in International Conference on Artificial Intelligence in Education, Springer, Cham, pp. 2-13. https://doi.org/10.1007/978-3-031-11644-5_1

Kontis, S. and Anastasiadou, S. (2026a) "Do Machines Feel?: The Emergence of Artificial Qualia and the Ethics of Emotion-Aware AI", in Seremeti, L., Anastasiadis, L. and Anastasiadou, S. (eds.), Philosophical Considerations of Computational Consciousness and AI Qualia, IGI Global Scientific Publishing, Hershey, PA, pp. 223-246. https://doi.org/10.4018/979-8-3373-5530-6.ch007

McStay, A. (2020) "Emotional AI and education: assessment of an electronic surveillance ecosystem", Policy Insights from the Behavioral and Brain Sciences, Vol. 7, No. 2, pp. 158-165. https://doi.org/10.1177/2372732220943222

Miao, F. and Holmes, W. (2023) Guidance for Generative AI in Education and Research, UNESCO Publishing, Paris.

Nussbaum, M.C. (2011) Creating Capabilities: The Human Development Approach, Harvard University Press, Cambridge, MA.

OECD (2025) Education for Human Flourishing, OECD Publishing, Paris.

Pekrun, R. (2006) "The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice", Educational Psychology Review, Vol. 18, No. 4, pp. 315-341. https://doi.org/10.1007/s10648-006-9029-9

Picard, R.W. (1997) Affective Computing, MIT Press, Cambridge, MA.

Ryan, R.M. and Deci, E.L. (2020) "Intrinsic and extrinsic motivation from a self-determination theory perspective", Contemporary Educational Psychology, Vol. 61, Article 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Sadek, M., Kallina, E., Bohné, T., Mougenot, C., Calvo, R.A. and Cave, S. (2024) "Challenges of responsible AI in practice: scoping review and recommended actions", AI & Society, Vol. 40, No. 1, pp. 45-62. https://doi.org/10.1007/s00146-024-01880-9

Seligman, M.E., Ernst, R.M., Gillham, J., Reivich, K. and Linkins, M. (2009) "Positive education: Positive psychology and classroom interventions", Oxford Review of Education, Vol. 35, No. 3, pp. 293-311. https://doi.org/10.1080/03054980902934563

Sen, A. (1999) Development as Freedom, Oxford University Press, Oxford.

Shneiderman, B. (2020) "Human-centered artificial intelligence: Reliable, safe & trustworthy", International Journal of Human-Computer Interaction, Vol. 36, No. 6, pp. 495-504. https://doi.org/10.1080/10447318.2020.1741118

UNESCO (2021) Recommendation on the Ethics of Artificial Intelligence, UNESCO Publishing, Paris.

Vistorte, A.Y.O., Deroncele-Acosta, A., Ayala, M., Barrasa, A., López-Granero, C. and Martí-González, M. (2024) "Integrating artificial intelligence to assess emotions in learning environments: a systematic literature review", Frontiers in Psychology, Vol. 15, Article 1122356. https://doi.org/10.3389/fpsyg.2024.1122356

Williamson, B. and Eynon, R. (2020) "Historical threads, datafication, and the governance of education", Learning, Media and Technology, Vol. 45, No. 3, pp. 223-235. https://doi.org/10.1080/17439884.2020.1772226