UNEC Tələbə Tədqiqatları Jurnalı, 3(1), 2026

DOİ: https://doi.org/10.30546/301678.01.010.2026.583

THE ALGORITHMIC PANOPTICON: EVALUATING TECHNICAL UNRELIABILITY, ETHICAL FAILURES AND THE REGULATORY FUTURE OF EMOTION AI IN EDUCATION

Nijat Ahmadov Shakir

Azerbaijan State University of Economics (UNEC), Digital Economics Faculty

E-mail: Ahmadov.Nijat.Shakir.2023@unec.edu.az

ABSTRACT

Emotion recognition AI is now present in education, with mixed and unintended consequences. This paper is focused on understanding what that introduction looks like specifically for neurodivergent pupils and those with diverse cultural backgrounds, for whom these technologies fare worst visibly. We will consider its particular limitations, significant ethical counter-criticism, and the need for strong non-supervisory arrangements that lag, focusing on the EU AI Act.

We also discuss the Act's Composition 5(1)(f) prohibits emotion recognition in educational settings as a general rule. Yet Annex III contemporaneously classifies certain functionalities, such as AI-driven gestures in exams as "high risk" instead of banned, which in turn impacts its transparency and accountability scores. In practice, neither provision has stopped deployment. Obtaining the student's consent is procedural at best, and not substantive. We have drafted an ethical integration framework and have pointed to empirically-based alternatives to biometric surveillance. The article concludes with recommendations of how to move away from the surveilled pedagogy to one of trust and ethics.

Keywords: *Emotion Recognition AI, Affective Computing, Educational Ethics, Student Privacy, Regulatory Compliance, Algorithmic Bias, Informed Consent, Pedagogical Integrity*

JEL classification codes: ***I21, O15, K23***

INTRODUCTION

In the realm of education, AI is often proclaimed as a means for personalized and adaptive learning. In the sphere of affective computing, the ERIAI is a tool in this future education model, and proponents of ERAI claim that a well-trained system can accurately forecast the students level of engagement, frustration, or boredom through a variety of biometric measures (Kapoor & Verma, 2024). However, in these celebratory depictions of technology, there is rarely a mention of the caveat of human emotions-the subjective and situational reality of human emotions-highlighted in

"Impossible AI" (Kumar, 2024), as in some AI systems there is simply not the ability to do what we are asking, the correlating of external, biometric measures of emotions to internal human emotion, and, in this sense, is an invasion into the classroom and classroom surveillance under the disguise of education.

Not only are there limitations in ERAI technology itself, but also in the implementation of ERAIs due to: which data is used to train ERAI, student/cultural differences in expressing emotions, effects of neurodivergence on ERAI systems. We will also address the validation gap, the discrepancy between a system with a high accuracy rate in a controlled lab setting and a lower accuracy rate in a real classroom setting. Although global effects of the EU AI Act will be addressed, the full scope of regulation cannot be understood without also exploring other types of intervention, such as "Smart Classroom Behavior Management" (Wang et al., 2023) in China and state surveillance within education.

To evaluate the consequences of ERAIs, it is important to distinguish the forms of ERAI, as technology is highly varied and spans from earlier Facial Expression Recognition (FER) systems to newer Multimodal Large Language Model (MLLM) based detection systems. While FER is based on the prediction of minute micro-expressions, the newer MLLMs, like AffectGPT, focus on the creation of "empathetic interaction" and thereby introduce concerns over student manipulation and student-teacher relationships (Lian et al., 2025; Chavan et al., 2025).

This paper will delve into the technical, ethical, and regulatory implications of ERAIs in education, explore the limitations and violations of consent within ERAI systems, and investigate varied regional regulatory frameworks, before a conclusion can be reached about ethical AI Governance and the alternative pedagogical practices.

Table 1. Evolution of emotion aI modalities and educational risks

AI modality	Biometric signal	Primary risk	Educational impact
Facial recognition	Micro-expressions	Cultural/neurodivergent bias	Misclassification of engagement
Vocal analysis	Pitch, tone, cadence	Linguistic/accent bias	Unfair assessment of participation
Physiological	HRV, skin conductance	Privacy intrusion	High surveillance anxiety
MLLMs (e.g. affectGPT)	Integrated multimodal data	"Empathetic" manipulation	Erosion of trust; shift from detection to behavioral influence

Source: *Table 1 provides an overview of ERAI mechanisms and their risks, informed by foundational work in FER* (Nagata & Okajima, 2024), *Vocal Analysis* (Kapoor & Verma, 2024), *and LMMs* (Ge et al., 2024).

Technical limitations

Cross-cultural validity and bias

A central observation in the most recent research literature is the cultural non-generalizability of facial expression recognition (FER) systems. FER systems that are trained on western faces performed very poorly when tested on eastern, african and other non-western faces; this violates the statistical principle of "measurement invariance". For example, the results showed that East Asians are more prone to attribute the same face to "fear" than “surprise”, while they clearly show significantly lower classification rates between "disgust", "anger", and "sadness" than westerners. Nevertheless, majority-label databases used for training FER systems completely ignore the differences in culture specific interpretation patterns (Jack et al., 2009, 2012). This is more than just a technical error in terms of accuracy; it is fundamentally a measurement problem: emotion inferences are simply not reliable across different cultures.

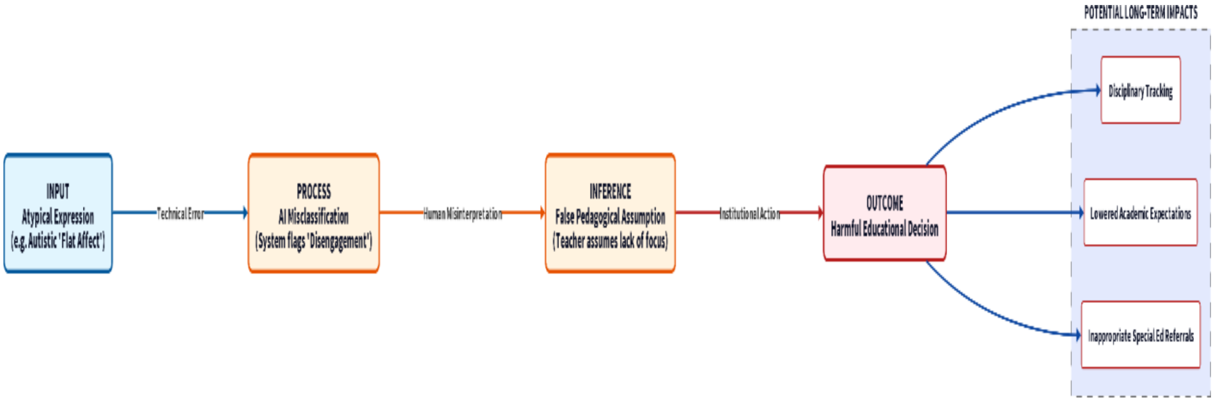
The measurement problem and neurodiversity gap

Beyond cultural bias, emotions are subjective and context dependent, which makes them difficult to categorize using fixed algorithms. When AI models are trained on standardized expressions like posed happy or sad faces, their accuracy drops significantly when analyzing spontaneous real-world expressions (Küntzler, 2021). This matters for neurodivergent students because their emotional expressions often diverge from the universal templates these systems are built on. Autistic individuals, for instance, frequently display facial expressions that are less extensive, more atypical, or otherwise distinct from neurotypical norms (Briot et al., 2021). As a result, AI systems trained predominantly on neurotypical data tend to misinterpret these expressions. Recent work with multimodal LLMs and foundation models has shown some improvement under controlled conditions. Deng et al. (2024) used the AV-ASD dataset to demonstrate that combining audio-visual and speech models detects autism-related behaviors more accurately than traditional facial expression analysis. Similarly, Zhang et al. (2025) introduced UniFER-7B, which enhanced both accuracy and interpretability for classification and also examined multimodal LLMs for facial expression recognition. These findings suggest that newer architectures could eventually support neurodivergent individuals more effectively (Sadigov et. al.,2025).

Yet most of these advances remain confined to research settings and specialized diagnostic tools. They are not yet general-purpose classroom monitoring systems. Even with improved accuracy, multimodal LLMs still face unresolved issues around reasoning and transparency (Zhang et al., 2025), which raises concerns if deployed for high-stakes educational decisions.

This figure depicts the cascade from atypical expressions in neurodivergent students through algorithmic misclassification to harmful institutional decisions.

Figure 1. The risk pipeline of emotion AI misclassification.



Source: *Conceptualized by the author based on ethical risks identified in Verhoef & Fosch-Villaronga (2023).*

This figure depicts the cascade from atypical expressions in neurodivergent students through algorithmic misclassification to harmful institutional decisions.

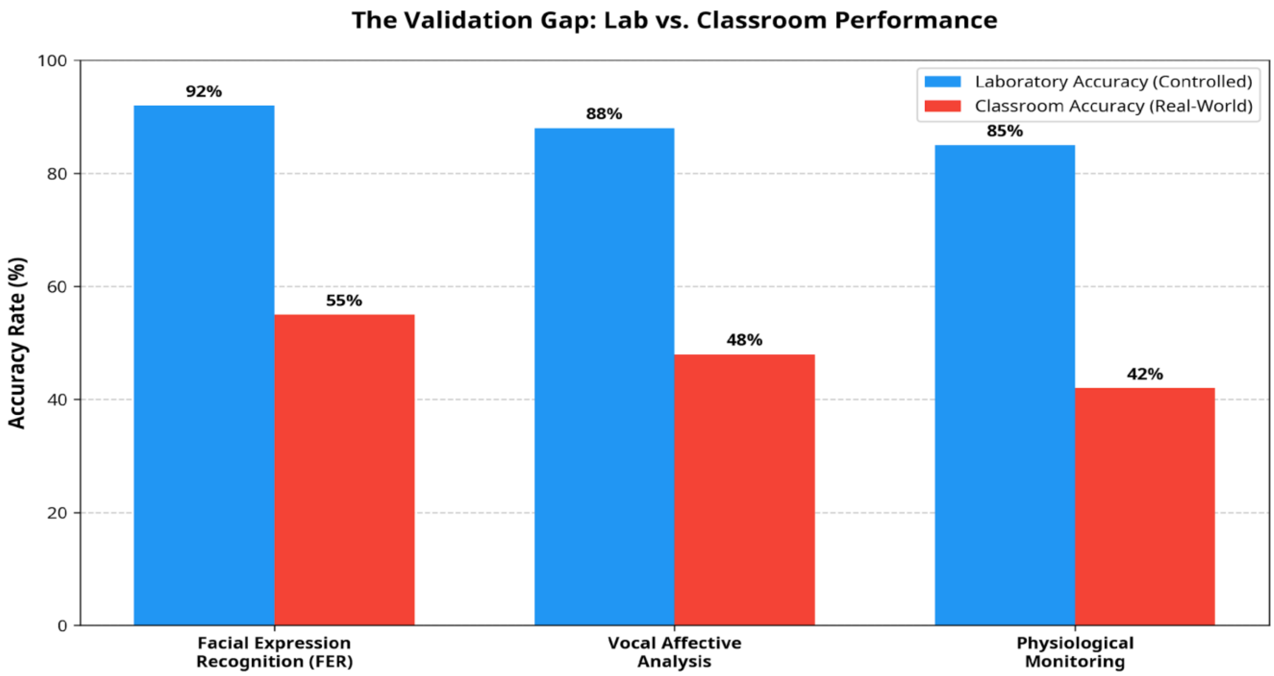
Some studies propose that emotion recognition technology might help autistic children practice social skills (Garcia-Garcia et al., 2022). Others caution that misclassification carries serious risks. Errors can lead to inappropriate disciplinary actions, incorrect special education placements, and lowered academic expectations, especially when a student's neutral or engaged expression is read as frustration or disinterest. The same measurement problems affect voice and physical signals, both of which vary by culture and neurotype.

The validation gap

Systematic reviews keep turning up the same problem. Researchers call it the "validation gap" or sometimes the "technology-education gap" (Shingjergji, İren, Urlings, & Klemke, 2026). One review looked at 96 studies on affective computing in online college education, published between 2019 and 2024. Most of those systems never made it out of the lab. They were tested in controlled settings, not in real, messy classrooms where the light changes, students fidget, and emotions show up spontaneously. So when you see high accuracy numbers, those probably won't hold up once you actually try to use the technology in practice.

One of the phase deployments throughout secondary and higher education was a duration of six months involving more than 300 students and 25 teachers (P. A. et al., 2025). A CNN-LSTM hybrid-based architecture was utilized for the learner performance prediction on a performance level with F1-score 0.91 and a multi-modal emotion recognition module using audio-visual based input. While performance of 88.3% for recognizing stress, confusion and satisfaction was attained in an environment based controlled situation, however, it has declined in a real classroom.

Figure 2. The validation gap in emotion AI



Source*: Illustrative ranges inferred from narrative synthesis of 96 reviewed studies (Shingjergji, İren, Urlings, & Klemke, 2026).*

Despite the fact that the pilot said that they increased the post-test results by 17%, engagement time by 28% and dropout rate by 22%, this is mainly because of the adaptive way content is delivered, since the tasks are delivered either with higher or gamified levels when the student is tired.

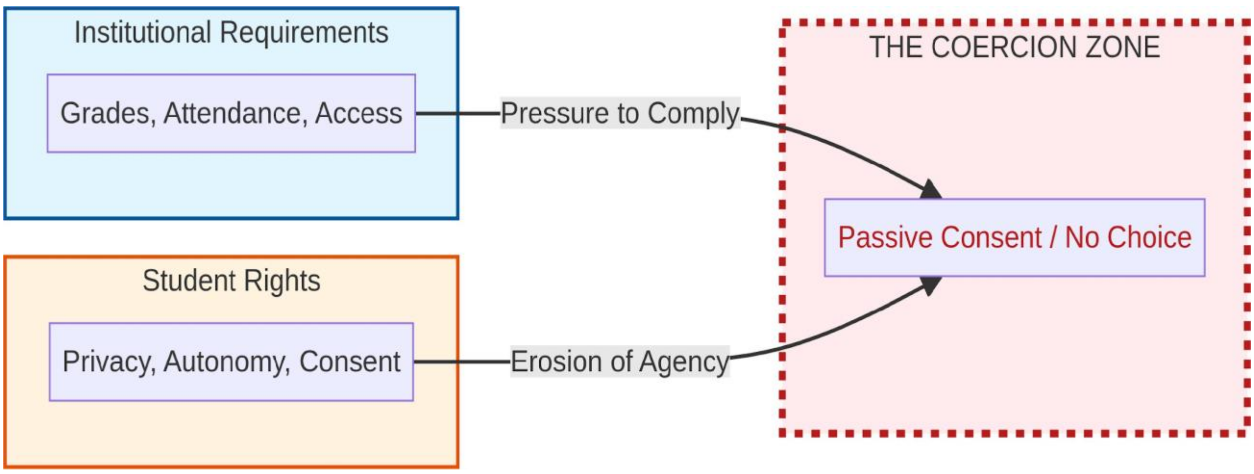
The system had better results because of the speed at which it responded to the student's performance, not due to the biometric inference layer. The override data are more difficult to brush aside. About 34% of the sessions had a manual override of the emotion detection module; a good fraction of this occurred with autistic students, as well as students from culturally varied backgrounds, as causes of misdetection. One third of the sessions is not a negligible error rate; this poses the straightforward question as to what is actually contributing to pedagogical intelligence in these implementations, and if the added component of emotion recognition-as opposed to passive indicators of behavior, like click sequences, time on task, and task completion-has been shown to be anything useful at all.

In another large observational study, Yuvaraj et al. (2025) used ensemble face and speech emotion recognition and LLM sentiment analysis on 367 classroom video snippets, obtaining a correlation $r = 0.513$ with human ratings of teacher warmth and supportiveness. This agreement was explained largely by features related to text sentiment, with face emotion recognition playing only a small role. They intended their study for purposes of generating average lesson-level climate for teachers, rather than monitoring specific students, and so this study offers little evidence for the ability of these kinds of systems to support high-stakes claims about particular students. A few constants seem to run through both papers. Performance drops off sharply in un-engineered settings like those where lighting can vary, students can move unexpectedly, and expressions are not acted out as prescribed. Real value add over and above current behavioral analytics capabilities of understanding emotional expression remain unsubstantiated. Teacher overriding during live operation really proves it is not being done as a compliance feature, but as part of a requirement. Organizations exploring these systems should ask for in-class operational results.

Ethical Implications

Educational institutions often implement emotion recognition artificial intelligence (ERAI) under conditions of “passive consent”, categorizing students as a “sensitive group” susceptible to cognitive-behavioral influence while lacking adequate protective measures (Gašević & Merceron, 2022). Students are commonly placed in a position of “consent-coercion”, where the sole practical alternative to biometric monitoring is to relinquish access to essential coursework, examinations, or institutional learning platforms (Pechenkina, 2023). This dynamic effectively establishes a coercive framework in which declining to participate in affective surveillance implicitly risks academic penalties. Empirical evidence highlights the tangible harms of such systems: Choksi et al. (2024) document that students under invasive proctoring and monitoring regimes have encountered baseless allegations of academic dishonesty, punitive disciplinary processes, and damage to their reputations. These outcomes reveal how surveillance infrastructure, even when formally presented as “optional”, can impose serious academic consequences on individuals who resist its implementation or deviate from its algorithmic expectations. Such a power asymmetry is inherent to educational environments and contravenes the core data protection principles of proportionality and minimization mandated by regulatory frameworks such as the General Data Protection Regulation (GDPR).

Figure 3. The Consent-Coercion Tension Model in Educational AI Deployment



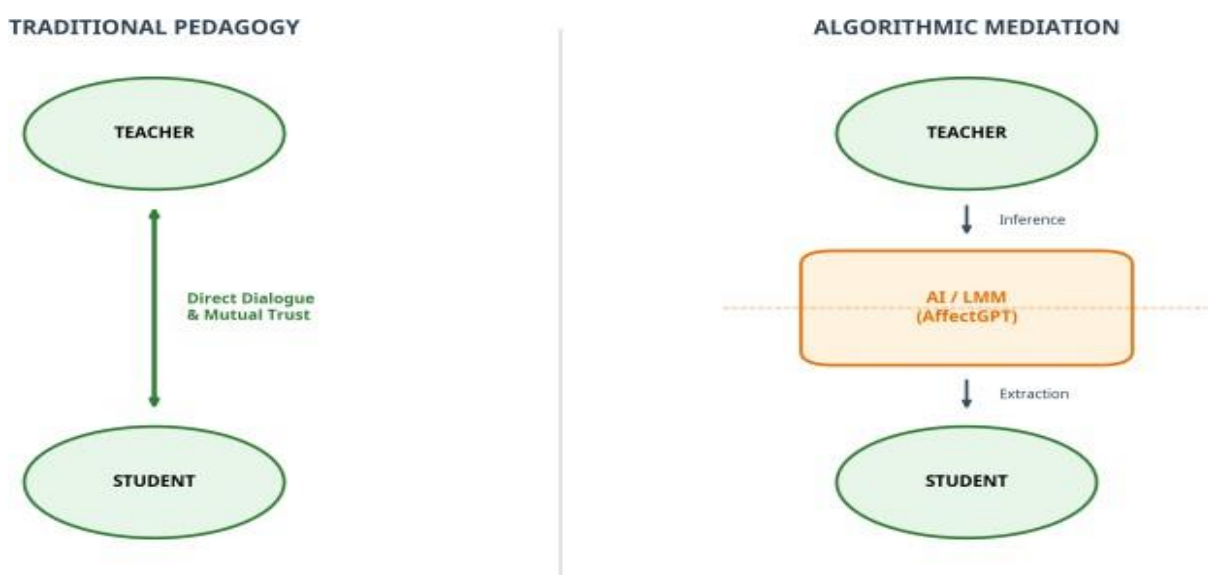
Source: *Figure is prepared by the author*

Student autonomy and the panopticon effect

The continued monitoring of emotions can lead to what researchers have termed the "panopticon effect", building on Foucault's concept of the structure of surveillance which implies that the awareness of being watched directly changes behaviour and limits personal freedom (Mörike, 2021). Students who are aware that their emotional display is being recorded and monitored may fail to express their emotions authentically, feel compelled to exhibit appropriate emotions, or feel anxious about misinterpretation of their emotions by the algorithms (Han et al., 2025). Such performative adaptations to monitoring in fact hinder true engagement and emotional display, thus impeding the educators' awareness of student affect (Han et al., 2025).

In addition, framing emotions as "data for extraction" and algorithmic scrutiny risks degrading the trust-dependent rapport between teacher and student, a cornerstone of successful teaching practice. As human interaction becomes filtered through computational interpretation instead of direct interpersonal exchange, the relationship can transition from being grounded in reciprocal comprehension and trust to one oriented around data harvesting and automated judgment.

Figure 4. The algorithmic mediation model vs. traditional pedagogy



Source: *Conceptualized by the author based on the "Panopticon Effect" and pedagogical trust theories (Mattioli & Cabitza, 2024).*

Recent progress in large multimodal models (LMMs), such as open-vocabulary frameworks like AffectGPT, reveals “emerged capabilities” that exceed the performance of earlier CNN-based architectures on standard benchmark tasks (Zhang et al., 2025). However, these technical improvements fail to address the foundational panopticon effect inherent to such systems. Whether an algorithm assigns a student's expression to a fixed category like "bored" or generates a more nuanced description such as "slightly disengaged, perhaps tired", the student is still placed under a regime of continuous affective surveillance. The core pedagogical harm persists regardless of the model's architectural sophistication. Consequently, instead of fostering teacher comprehension through direct interaction and dialogue, student understanding becomes filtered through systems prone to misclassifying emotional states. This risk is especially pronounced for culturally diverse and neurodivergent student populations (Mattioli & Cabitza, 2024).

Regulatory landscape and the EU AI act

The EU AI Act isn't merely a recommendation of good manners, then. It's a fairly hard red line for schools and universities. Article 5(1)(f) states a key prohibition: "prohibited use of systems for emotion recognition in educational institutions". However, as always, the nuances are hidden in the definitions:

To clarify what is meant by this prohibition, one must consider Article 3(39). The AI Act defines an emotion recognition system as "any AI system intended to identify or infer one or more of a person’s biometric data, one or more emotional states or the personality traits or features of a person...". The EU Commission has specified in February 2025 that this means that emotion recognition is only prohibited when the related system deals with biometrics (e.g. Facial expressions, heartbeat or fingerprint). This is crucial for the ed-tech industry, as any sentiment analysis tools that operate only based on text analysis are usually allowed as they do not require biometric data.

However, there is also a small exception for "medical or safety reasons", but do not count on it to apply for any general "student wellness" app; the Commission is very clear that this only covers legitimate medical devices, not a sentiment analysis tool for the entire classroom.

The gray area: monitoring behavior vs. reading minds

The distinction between forbidding emotion reading and allowing behavior monitoring is unclear. While you can't try to "read" a student's inner feelings, Annex III(3)(d) actually allows AI systems that monitor "prohibited behavior" during exams, basically, AI proctoring. Proctoring instruments are marked as "high-risk" not as "prohibited."They can be deployed legally, but in return a host of additional homework will follow, such as human review (Article 14) and human rights impact assessments (Article 27). The Commission attempts to distinguish between the two by defining monitoring eye movement or head position as "behavior", and attempting to infer a students' frustration as "emotion".

But out in the wild, the line feels a little hazy. A system designed to track a student's eyes in order to detect cheating could be repurposed just as easily to measure engagement levels. The Commission is aware of the possibility of companies attempting to rebrand the technology and use terms like "behavior analytics" instead, allowing them to get around the ban. With regard to schools this will make it important to look beyond the terms companies use and inspect what the technology is actually doing.

Providers vs. deployers

On top of the ban and high-risk tiers, article 50(3) provides for transparency obligations for deployers of emotion recognition systems. It states that "the deployer informs the natural persons that they are being exposed to an emotion recognition system". This obligation applies irrespective of the system falling in the banned category (as it should not be deployed then) or in the high-risk category, it reiterates the intention of the Act that emotional surveillance never happens without the subjects being warned. Breach of Article 5(1)(f) is punishable by the highest penalties stipulated in the Act (up to 7% of worldwide annual turnover (Article 99)).Amongst the confusion of this complex regulatory structure, only one dichotomy is critically important for schools and universities attempting to understand it: between a "provider" and a "deployer". According to Article 3(3) of the regulation, a "provider" is a natural or legal person that has developed an AI system or has commissioned the development of that AI system, and places it on the market or puts it into service under its own name. Under Article 3(4) the "deployer" means an entity which uses an AI system under its own authority, two separate categories and the list of what they are responsible for; incorrectly using the terms for the purchase of Ed Tech, the misuse could distinguish a correct and compliant usage, or incorrect and liable usage.

Provider obligations for high-risk and prohibited systems

For the providers of high-risk AI systems, Article 16 imposes an obligation to ensure compliance with the provisions set out in Chapter III Section 2 (Articles 8-15): risk management system (Article 9), data governance (Article 10), technical documentation (Article 11), record keeping (Article 12), transparency and information (Article 13), human oversight (Article 14) and accuracy, robustness and cybersecurity (Article 15). Providers are additionally obliged to adopt a QMS (Article 17), to retain documents for ten years (Article 18) and logs that are automatically produced (Article 19), to undertake conformity assessment (Article 43), to draw up the EU declaration of conformity (Article 47), to affix the CE mark (Article 48) and to register the AI system in the EU database (Article 49). In this one sense these obligations for systems to identify emotions within education are irrelevant: placing these systems onto the market for education is prohibited according to Article 5(1)(f). A way for providers to get around this would be to market their system as an "analytics tool" or a "safety tool", leaving the burden to the deployer to evaluate if the function of the AI system really does involve the prohibited emotion recognition.

Deployer obligations and institutional compliance

Specifically,Article 26places several obligations on deployers of high-risk AI systems, including: (a) putting in place appropriate technical and organisational measures to ensure operation in line with instructions, (b) having in place human oversight capabilities held by able and competent persons with the authority to intervene, (c) ensuring the input data is relevant and representative, (d) carrying out operation of the system and reporting of serious incidents to the provider and to the market surveillance authority, and (e) retaining logs for a minimum of six months. Article 27 additionally obliges deployers to conduct a Fundamental Rights Impact Assessment (FRIA) before an high-risk AI system is put into use. For educational institutions, this implies that a behaviour monitoring tool (allowed under Annex III(3)(d)) cannot be deployed without demonstrating available human oversight, having trained staff available for the use of the system, and conducting an assessment on students' right to privacy, dignity, and equal treatment. UnderArticle 25(1), a"deployer-becomes-provider" mechanism further ensures that if a substantial modification to a high-risk AI system, or a modification to its intended purpose that creates a high-risk use, is made by an educational institution, the institution becomes responsible as if it were the provider of the AI system and is therefore subject to conformity assessment and CE marking requirements with which its IT department is likely to be unable to comply, as will be further discussed with respect to a customized emotion inference system.

The internal deployment consideration

There is also a key uncertainty concerning internally developed systems. Pursuant to Article 3(3), a university that develops an emotion recognition system for internal use and no public marketing of it will not qualify as a 'provider'. Recital 82, however, states that the purpose of the Act extends to situations regardless of market availability of a system. A university will still fall under a violation of Article 5(1)(f) if it places into operation a custom system to conduct impermissible emotion inferences. 'Putting into service' covers the case of a university deploying an in-house built system, so universities and schools cannot sidestep the prohibitions through custom developments. A tool that is designed in-house to continuously monitor high-risk behaviour can hardly avoid making a provider default Article 16 obligation, if it has no other providers whom to entrust it to carry out provider duties. This produces a discrepancy: the technical conformity duties of a procurement of a commercial tool bearing CE marking will be with its vendor, but that responsibility will be with the institution if it has developed its tool.

Strategic framework for ethical AI integration

In order to deal with the issues listed above, we have devised a framework to help educational institutions plan and adapt the integration of AI.

Phase 1: Ethical scoping and impact assessment

Data audit: Assess whether the proposed AI tool has been validated for the specific demographic of the student body (race and neurodiversity). Under a Neurodiversity-affirming paradigm, validation must explicitly reject tools designed for "masking" or "normalization" training, approaches that pressure autistic and neurodivergent students to suppress authentic expressive patterns in order to conform to neurotypical algorithmic norms (Briot et al., 2021). Instead, the audit should verify that the system respects diverse expressive styles as valid rather than treating neurotypical patterns as the universal standard to which all students must adapt.

- **Necessity test:** Determine if the educational goal can be achieved through less intrusive, non-biometric means.
- **Risk mapping:** Identify potential downstream effects, such as biased discipline referrals.

Phase 2: Transparency and meaningful consent

- **Opt-in methods:** Shift from passive to a more affirmative, explicit and periodic form of "adaptive consent".
- **Explainability:** Provide transparency to teachers and students about the rationale behind all AI-driven inferences.
- **Human-in-the-loop:** Enforce a rule where automatic inferences do not lead to a disciplinary or academic consequence unless a human is included in the loop.

Phase 3: Ongoing auditing and governance

- **Auditing for bias:** Regularly perform technical audits for measurement invariance across all student subgroups.
- **Data minimization:** Adopt a strict data deletion protocol for affective data once no further instructional utility exists.
- **Pedagogical framing:** Ensure that AI complements, not supplants, the teacher-student trust relationship.

Evidence-based alternatives

The regulatory shift and technical unreliability of emotion recognition necessitate exploration of evidence-based alternatives that achieve educational goals without intrusive biometric surveillance (August & Tsaima, 2021). These alternatives prioritize pedagogical integrity and student autonomy while supporting student engagement and emotional well-being.

1. **Direct communication**

Instead of using algorithms to guess a students' "frustration" or "confusion", we can talk directly to them to understand their learning experience. This is "human-in-the-loop" learning in that educational professionals, instead of extracting data to find what response the students may need, will directly communicate with the student to find out what response is needed. Direct communication allows for more pedagogical benefit to the learning experience because it does not overstep student agency, provides correct data about the student's needs, and strengthens the trust relationship between the student and teacher.

2. Observable behavioral indicators

Institutions can assess engagement through non-invasive, objective data points including attendance patterns, assignment completion rates, participation in forums or live sessions, and time spent on learning activities. These metrics do not require inference of internal emotional states and thus avoid the measurement problem associated with emotion recognition. Observable behavioral data, such as assignment completion rates, participation frequency, and time-on-task metrics reflects actual student actions rather than algorithmic interpretations of facial expressions or voice patterns, making it less susceptible to the cross-cultural and neurodivergent misclassification errors that undermine Facial Expression Recognition systems (Jack et al., 2009; Verhoef & Fosch-Villaronga, 2023).

3. Formative and participatory assessment

Biometric data is not always required to get instant data on student learning. Low-stakes formative assessments, such as think-pair-share activities, quick comprehension checks, or reflective writing provide immediate feedback on student understanding without biometric monitoring. These assessment approaches are pedagogically sound, engage students in metacognition (Cox et al., 2022), and provide educators with direct evidence of learning rather than inferred emotional states.

Participatory assessment approaches that involve students in evaluating their own learning further strengthen student agency and autonomy (Nieminen et al., 2025).

4. Student self-report and voluntary sharing

Instead of automatically inferring emotions, institutions can move to consensual reporting, where students express their feelings in self-surveys, reflection journals, or conversations with their teacher. This respects students' privacy and agency, while providing teachers with students' own description of how they feel. The students are telling the teacher how they are feeling and how they are learning, and getting their feedback is the only ethical way to ask how they are feeling (Banzon et al., 2023).

Human oversight testing and compliance operationalization

Far from asking for human supervision, the AI act requires it. Articles 14 and 26 require: "effective oversight by natural persons, education, training and authorization of personnel involved in high-risk AI systems", with adequate resources to support those working in such areas. This is a useful protection, but the EU AI Act is remarkably silent about what we need to do in order to test the effective human oversight in the classroom. It is likely that schools and universities will fall into the compliance trap: ticking boxes while the protection of the students remains unaddressed. A Three-Tier Oversight Evaluation Framework. Drawing on the Metrics-Driven Human Oversight Framework (Kandikatla, L. (2026)) and the conformity assessment structure of Annex VI, we propose a tiered approach for testing oversight efficacy in educational AI deployment, adapted to the risk tier of the system in question.

Tier 1: Human-in-the-Command (HIC)

All high risk AI systems falling under the scope of Annex III(3)(d) (and thus system monitoring and proctoring tools) shall have an authorized member of senior administration/committee assigned responsible for suspension of the AI system operation. This category demands clear governance structure stating (a) situations under which system shall be suspended (e.g. System failure resulting in over threshold discrimination, complaint by a student, request from a regulatory body), (b) the authority and timing of decision on the suspension and (c) the responsibility when failure in monitoring system results in discrimination. Approach to testing is tabletop exercise replicating false positives (neurodivergent students are erroneously over-represented as system failure on system while in a high-stakes testing situation). How quickly the nominated review body is able to execute the suspension process in under 24 hours.

Tier 2: Human-in-the-Loop (HITL)

Where AI output is constant and has implications for pedagogical decision-making (e.g. if engagement analysis or attention scores are monitored and then affect the pedagogical action a teacher takes), under Article 14 this requires human involvement of a suitably trained teacher, before the predictions affect a child's grades, or disciplinary actions. HITL requires: a) dashboards for presenting output with confidence intervals to a human; b) a technologically-simple and operationally-unimpeded process for overrides; and c) explicit intervention logging-tracking exactly when and for what reasons an educator overrides the algorithm. We will look at override rates and time-to-override in simulation, and derive benchmarks indicating that at least 95% of all algorithmic alerts need teacher involvement before reaching students. As mentioned above, the manual override percentage in the previous system was 34%. This was too high for teachers to keep up, to retain control over what reaches students if too many false positives are raised, thus not achieving "effective oversight".

Tier 3: Human-on-the-Loop (HOTL)

For the low-risk analytics (i.e. approximating classroom climate), direct monitoring might not be needed, and an audit by the responsible party would be sufficient. The Article 26(1) sentence that 'the deployer of a high-risk system shall monitor its operation' means: a) the system performance will be monitored every 6-months through automatic reports that are checking on the demographic parity of the system outputs; b) a monthly report about the instructor satisfaction on how the quality of the system has improved will be generated by a monthly process; and c) third party audit on the policy on deleting/keeping data every 3-months to make sure that the policy described on Article 17 'data minimization' is maintained. For testing, KPI's will be given for each testing phase, while reporting all problems to HITL/HIC so any abnormality could be found (Teymurov et al., 2026).

Conformity assessment and documentation

Under Article 43, high-risk AI systems must go through a conformity assessment procedure before they may be placed on the market and will require self-assessment under internal controls (Annex VI) or third-party examination by a notified body. For educational institutions, the following elements are therefore needed for behavior monitoring tools: (a) technical documentation (including a description of the system's risk management system - Article 9, specification of its training data-Article 10, description of the interface where human oversight will be conducted - Article 13(3)(d)); (b) instructions for use, including any intended educational use, known risks, and reasonably foreseeable misuse; (c) an EU declaration of conformity (Article 47) to be kept for ten years; The institutions should request vendors to supply the above as a condition of purchase and should check CE marking through the EU database (Article 49) prior to implementing the system. For institutions who plan to develop a system in-house, they will need to produce the technical documentation themselves, which should be weighed against the decision of to buy or to build.

Post-market monitoring and incident reporting

In order to comply with Article 72 providers should implement post market surveillance systems that, during the entire lifecycle of the system, "actively and systematically collect, store and analyze relevant information". The participation of the deployer in this surveillance is based on notifying providers and market surveillance authorities in case of an important incident (Article 26(5)). In an academic environment the market surveillance is based on the existence of 3 factors: (a) an external third-party audit of bias conducted annually; the report of the audit must also be used to assess whether the accuracy of results varies among demographic groups; (b) student voice mechanisms (surveys, focus groups, student committees); feedback from students on intrusiveness and pedagogical utility should be sought; (c) reporting systems to ensure the provider is notified when results are discriminatory or when there are technical faults. Post-market surveillance thus takes on the dimension of a governance process in this context that is not limited to regulatory obligations and is part of Article 26 deployer duties: protection of the students extends beyond the minimum prescribed.

CONCLUSION

Emotion Recognition AI (ERAI) in education is a risky business concerning student's private information and freedom of action. Personalized learning has a good-sounding ring to it, but present-day ERAI systems are not dependable to the extent we should trust it with so much-also the technology has issues with ethics. Pilot tests on the application of these systems show that these are not highly accurate systems. For instance, students are often misclassified-especially neurodivergent and non-white students. Thus, they are so flawed that teachers constantly have to intervene to repair its classification, so the technology can become a distraction rather than assistance.

While the EU AI Act seeks to curtail these practices, it does so through a rather intricate piece of legislation. Article 5(1)(f) completely bans generalized AI emotion recognition in schools for being "unacceptable risk" yet "high risk" carve-out is provided in Annex III(3)(d) to AI which "monitors pupils' behavior when they are sitting examinations", provided that certain conditions such as on transparency and human control, are fulfilled. The vagueness on what is disallowed, what is permitted, therefore, needs to be tackled with careful case by case consideration and responsibility being taken by the schools, irrespective of whether the tools used are bought or developed in-house, to take responsibility for effective supervision, which entails the requirement for significant human involvement, fundamental rights analysis, and ongoing monitoring. The three-tier supervision system developed by us can be a tangible way of doing more than "lip service" for protecting students' well-being. There is significant variation in the regulatory landscape throughout the world. Although the EU has established an effective blockade within the classroom, China is rapidly expanding "Smart Classroom Behavior Management", a system that never ceases to observe the biometrics of its pupils. Comparing the two displays that no blind assumption can be made that AI within the classroom is not a moral and political battlefield and within the education system, it is the political actor's stance of resistance to algorithmic power that matters. It is those tools which foster belief in the student, and not those which permit an omniscient surveillance system, that the political leader within education worldwide should prioritize.

REFERENCES

1. August S. E., Tsaima A. (2021). "Artificial intelligence and machine learning: An instructor's exoskeleton in the future of education". In: Ryoo J., Winkelmann K. (Eds.), *Innovative learning environments in STEM higher education*. SpringerBriefs in Statistics. Springer, Cham. https://doi.org/10.1007/978-3-030-5898-6_5

2. Banzon A. M., Beever J., Taub M. (2024). "Facial expression recognition in classrooms: Ethical considerations and proposed guidelines for affect detection in educational settings". *IEEE Transactions on Affective Computing*, Vol. 15, No. 1, pp. 93-104. https://doi.org/10.1109/TAFFC.2023.3275624

3. Briot K. və b. (2021). "New technologies as promising tools for assessing facial emotion expressions impairments in ASD: A systematic review". *Frontiers in Psychiatry*, Vol. 12, Article 634756. https://doi.org/10.3389/fpsyt.2021.634756

4. Chavan V. və b. (2025). "Feeling machines: Ethics, culture, and the rise of emotional AI". arXiv. https://doi.org/10.48550/arXiv.2506.12437

5. Choksi M. Z. və b. (2024). "Student privacy activism: Student-organized activism and obfuscation respond to intrusive surveillance in digital assessment within higher education". *Journal of Information Policy*, Vol. 14. https://doi.org/10.5325/jinfopoli.14.2024.0017

6. Cox S. M., Jongbloed K., Black C. (2022). "Metacognition in teaching: Using a 'rapid responses to learning' process to reflect on and improve pedagogy". *Teaching & Learning Inquiry*, Vol. 10. https://doi.org/10.20343/teachlearninqu.10.27

7. Deng S. və b. (2025). "Hear me, see me, understand me: Audio-visual autism behavior recognition". *IEEE Transactions on Multimedia*, Vol. 27, pp. 2335-2346. https://doi.org/10.1109/TMM.2024.3521838

8. European Commission (2025). *Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)*, C(2025) 884 final. https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act

9. European Parliament and Council of the European Union (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union. http://data.europa.eu/eli/reg/2024/1689/oj

10. Future of Privacy Forum (2025). *Conformity assessments under the EU AI Act: A practical guide*. https://fpf.org/fpf-resources-on-the-eu-ai-act/

11. Garcia-Garcia J. M. və b. (2022). "Using emotion recognition technologies to teach children with autism spectrum disorder how to identify and express emotions". *Universal Access in the Information Society*, Vol. 21, pp. 809-825. https://doi.org/10.1007/s10209-021-00818-y

12. Ge M., Tang D., Li M. (2024). "Video emotion open-vocabulary recognition based on multimodal large language model". arXiv. https://doi.org/10.48550/arXiv.2408.11286

13. Han J. və b. (2025). "Pioneering multimodal emotion recognition in the era of large models: From closed sets to open vocabularies". arXiv. https://doi.org/10.48550/arXiv.2512.20938

14. Jack R. E. və b. (2009). "Cultural confusions show that facial expressions are not universal". *Current Biology*, Vol. 19, pp. 1543-1548. https://doi.org/10.1016/j.cub.2009.07.051

15. Jack R. E. və b. (2012). "Internal representations reveal cultural diversity in expectations of facial expressions of emotion". *Journal of Experimental Psychology: General*, Vol. 141, No. 1, pp. 19-25. https://doi.org/10.1037/a0023463

16. Kapoor A., Verma D. (2024). "Emotion AI: Understanding emotions through artificial intelligence". *International Journal of Engineering Science and Humanities*. https://doi.org/10.62904/0vcbvb24

17. Kandikatla L. (2026). "Metrics Driven Human Oversight Framework for AI Systems". Preprints. https://doi.org/10.20944/preprints202602.0374.v1

18. Kumar S. (2024). "To Be, Or Not To Be?: Regulating Impossible AI in the United States". arXiv. https://doi.org/10.48550/arXiv.2408.01440

19. Küntzler T. və b. (2021). "Automatic Facial Expression Recognition in Standardized and Non-standardized Emotional Expressions". *Frontiers in Psychology*, Vol. 12, Article 627561. https://doi.org/10.3389/fpsyg.2021.627561

20. Lang C. və b. (2022). *Handbook of Learning Analytics* (2nd ed.). Society for Learning Analytics Research (SoLAR). https://doi.org/10.1808/hla22

21. Lian Z. və b. (2025). "AffectGPT: A New Dataset, Model, and Benchmark for Emotion Understanding with Multimodal Large Language Models". In: *Proceedings of ICML 2025*. https://proceedings.mlr.press/v267/lian25a.html

22. Mattioli M., Cabitza F. (2024). "Not in my face: Challenges and ethical considerations in automatic face emotion recognition technology". *Machine Learning and Knowledge Extraction*, Vol. 6, No. 4, pp. 2201-2231. https://doi.org/10.3390/make6040109

23. Nagata M., Okajima K. (2024). “Effect of observer's cultural background and masking condition of target face on facial expression recognition for machine-learning dataset”. *PLOS ONE*, Vol. 19, No. 10, e0313029. https://doi.org/10.1371/journal.pone.0313029

24. Nieminen J. H. və b. (2025). “Student partnership in assessment: What works, for whom, why, and under what circumstances?”. *Review of Education*, Vol. 13, e70059. https://doi.org/10.1002/rev3.70059

25. Pechenkina E. (2023). “Artificial intelligence for good? Challenges and possibilities of AI in higher education from a data justice perspective”. In: Czerniewicz L., Cronin C. (Eds.), *Higher education for good: Teaching and learning futures*, pp. 239-266. Open Book Publishers. https://doi.org/10.11647/obp.0363.09

26. Sadigov R. A. və b. (2025). “Modern research of using alternative energy resources in Azerbaijan”. *International Journal of Advances in Applied Sciences*, Vol. 14, No. 3, pp. 907-915. https://doi.org/10.11591/ijaas.v14.i3.pp907-915

27. Shingjergji K. və b. (2026). “Affective computing in online higher education: A systematic literature review”. *Computers and Education: Artificial Intelligence*, Vol. 10, 100499. https://doi.org/10.1016/j.caeai.2025.100499

28. Teymurov M. A. və b. (2026). “Assessment of the synergistic impact of climate change and human activities on landscape ecozones”. *SABRAO Journal of Breeding and Genetics*, Vol. 58, No. 3.

29. Verhoef T., Fosch-Villaronga E. (2023). “Towards affective computing that works for everyone”. arXiv. https://doi.org/10.48550/arXiv.2309.10780

30. Wang Z. və b. (2023). “Student Learning Behavior Recognition Incorporating Data Augmentation with Learning Feature Representation in Smart Classrooms”. *Sensors*, Vol. 23, No. 19, 8190. https://doi.org/10.3390/s23198190

31. Yuvaraj R. və b. (2025). “Affective Computing for Learning in Education: A Systematic Review and Bibliometric Analysis”. *Education Sciences*, Vol. 15, No. 1, 65. https://doi.org/10.3390/educsci15010065

32. Zhang F. və b. (2025). “Rethinking facial expression recognition in the era of multimodal large language models: Benchmark, datasets, and beyond”. arXiv. https://doi.org/10.48550/arXiv.2511.00389

XÜLASƏ

Emosiyaların tanınması üzrə Süni İntellekt texnologiyaları artıq təhsil sisteminə daxil olub və bu proses bir sıra ziddiyyətli, eləcə də gözlənilməz nəticələrə yol açmaqdadır. Bu məqalə, sözügedən texnologiyaların tətbiqinin xüsusilə neyromüxtəlifliyi olan (neurodivergent) və müxtəlif mədəni mənşəyə malik şagird və tələbələr üçün hansı xüsusiyyətlərə malik olduğunu araşdırmağa yönəlmişdir, çünki bu qruplar üçün qeyd olunan texnologiyaların fəaliyyəti ən qeyri-dəqiq və mənfi nəticələrlə müşahidə olunur. Tədqiqatda Avropa İttifaqının Süni İntellekt haqqında Aktına (EU AI Act) diqqət yetirilərək, texnologiyanın spesifik məhdudiyyətləri, əhəmiyyətli etik tənqidlər və hazırda geridə qalan güclü qeyri-nəzarət mexanizmlərinə olan ehtiyac nəzərdən keçirilir.

Məqalədə həmçinin Aktın 5(1)(f) maddəsinin təhsil müəssisələrində emosiyaların tanınmasını ümumi qayda olaraq qadağan etməsi müzakirə olunur. Bununla belə, III əlavədə eyni zamanda imtahanlarda Sİ vasitəsilə jestlərin təhlili kimi müəyyən funksionallıqlar "qadağan olunmuş" deyil, "yüksək riskli" kimi təsnif edilir ki, bu da öz növbəsində şəffaflıq və hesabatlılıq göstəricilərinə mənfi təsir göstərir. Praktikada nəzərdə tutulan müddəaların heç biri bu texnologiyaların tətbiqini dayandırmamışdır. Tələbələrin razılığının alınması isə mahiyyət etibarilə deyil, ən yaxşı halda yalnız formal prosedur xarakteri daşıyır. Tərəfimizdən etik inteqrasiya çərçivəsi hazırlanmış və biometrik müşahidəyə dair empirik əsaslı alternativlər təklif edilmişdir. Məqalə "müşahidə edilən" pedaqogikadan etimad və etikaya əsaslanan pedaqogikaya keçid yollarına dair tövsiyələrlə yekunlaşır.

Açar sözlər: *Emosiyaların Tanınması üzrə Sİ, Affektiv Hesablama, Təhsil Etikası, Tələbə Məxfiliyi, Tənzimləyici Uyğunluq, Alqoritmik Qərəzlilik, Məlumatlı Razılıq, Pedaqoji Bütövlük*

JEL kodları: ***I21, O15, K23***