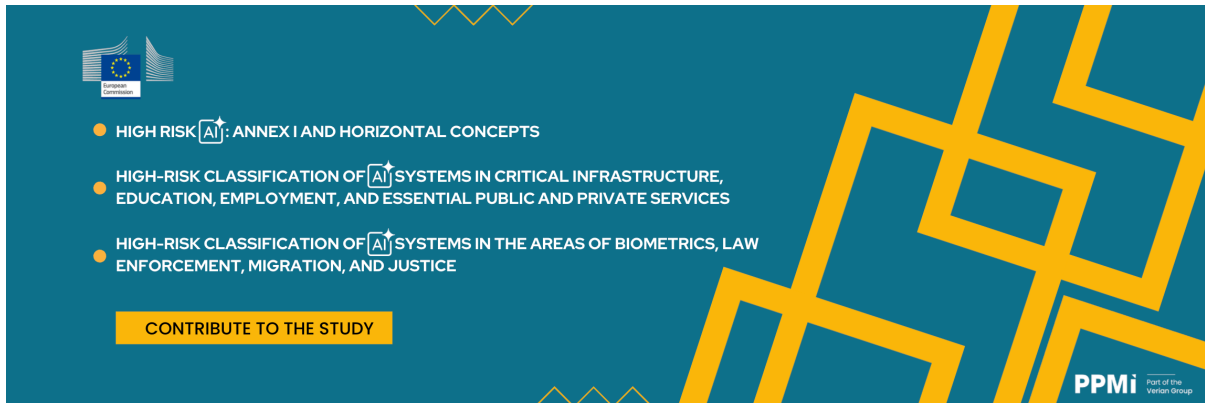


High Risk AI: Feedback Collection

Fields marked with * are mandatory.



Thank you for your interest in contributing to this study.

This survey provides **an additional opportunity for stakeholders** who were unable to participate in the workshops to share their views.

The survey covers the same **three thematic areas** addressed during the workshops, namely:

- (i) **Annex I and horizontal aspects** related to the high-risk classification of AI systems;
- (ii) High-risk classification of AI systems in **critical infrastructure, education, employment, and essential public and private services**; and
- (iii) High-risk classification of AI systems in the areas of **biometrics, law enforcement, migration, and justice**.

Respondents are free to **contribute to one, several or all topics**, depending on their expertise and interest.

The inputs collected through this survey will **complement the evidence gathered during the workshop discussions and feed into the overall analysis**. The survey focuses on practical experiences, challenges and points for clarification related to the application of the existing framework. **It does not seek feedback on the provisions of the AI Act itself**, which is not subject to discussion or amendment in the context of this study.

The survey is **open until 7 January**.

We thank you for your interest and for taking the time to share your insights.

* Full Name

Amina Fazlullah

* Your organisation

Common Sense Media

* Your role in the organisation

Head of Tech Policy Advocacy

* Your country

US - United States of America

* What topic would you like to contribute to?

- ☐ Annex I and horizontal aspects
- ☒ High-risk classification of AI systems in critical infrastructure, education, employment and essential public and private services
- ☐ High-risk classification of AI systems for biometrics, law enforcement, migration and justice

* What aspects of **High-risk classification of AI systems in critical infrastructure, education, employment and essential public and private services** would you like to cover?

- ☐ Critical infrastructure
- ☒ Education and vocational training
- ☐ Employment, workers' management and access to self-employment
- ☐ Essential public and private services and benefits
- ☐ Other

High-risk classification of AI systems in critical infrastructure, education, employment and essential public and private services

Education and vocational training

1. Contractual arrangements for AI system deployment: Are you aware of common practices or contractual arrangements where educational or Vocational Education and Training (VET) institutions purchase AI systems

and deploy them under their own name or trademark and not under the trademark of the provider who provided the system? Under the AI Act, this would imply that the institution, as the deployer, assumes all obligations of a provider.

While overt white-labeling arrangements are not yet widespread, there is a significant and growing gray area that requires clarity. As AI capabilities become increasingly integrated into schools' existing technology stacks, the lines of responsibility and control are becoming blurred in ways that may create unintended regulatory gaps. Examples of this gray area include:

Learning Management Systems (LMS), which may be white-labeled, that have added AI features (auto-grading, content recommendations, plagiarism detection)

Student information systems that incorporate AI-powered analytics as part of regular updates

EdTech platforms adding AI tutoring or assessment features that schools didn't explicitly procure as separate AI systems

Third-party integrations and plugins that add AI functionality to existing school platforms

In these scenarios, it's unclear whether:

Schools understand that AI systems are being deployed within their technology infrastructure

Schools have sufficient technical access and documentation to fulfill provider obligations

The original software provider or the AI component developer bears responsibility

Schools have the technical capacity to conduct conformity assessments or risk management for AI features embedded in broader platforms

This gray area risks creating a compliance gap where neither schools nor vendors clearly understand who holds provider obligations under the AI Act. Educational institutions need clear guidance on when integration of AI into existing systems triggers provider obligations, what level of control/customization shifts responsibility to the deploying institution, and how to assess whether vendors are fulfilling AI Act requirements for embedded AI features.

Without clarity, schools may unknowingly assume provider-level obligations they cannot fulfill, or alternatively, AI systems may be deployed in education without adequate oversight because all parties assume someone else is responsible.

2. Processing of personal data and profiling: In your experience and in light of the use cases in point 3 of Annex III, can you share examples where the AI system would not involve profiling and could fall under one of the conditions for the exception in the filter in Article 6(3) AI Act?

3. Differences between Annex III, point 3(a) and point 3(c): What relevant differences do you see between AI systems covered by Annex III, point 3(a) and point 3(c) of the AI Act? How do these differences impact the application of the regulation?

The primary distinction is timing and purpose: 3(a) determines who gets in and where they're placed, while 3(c) evaluates how students perform once enrolled.

This distinction creates several practical challenges:

Many AI systems span both categories. For example, an adaptive learning platform that evaluates performance (3c) may generate data that subsequently informs placement in advanced or remedial programs (3a). It's unclear whether such systems should be classified under both provisions or how to separate these functions for compliance purposes.

Evaluation data from 3(c) systems may feed into 3(a) decisions at later stages. The regulatory framework needs clarity on how to handle these interconnected uses and whether downstream use of evaluation data for gatekeeping decisions triggers additional obligations.

Assessment systems initially deployed for formative evaluation (3c) frequently evolve to inform higher-stakes decisions about tracking, grade advancement, or program eligibility (3a). This raises questions about when obligations change and whether systems require re-classification.

While both are high-risk, the specific harms and mitigation strategies may differ. Schools and vendors need guidance on whether risk management and conformity assessment procedures should differ based on classification, or if systems serving both functions require comprehensive assessment across both use cases.

The educational sector would benefit from explicit guidance on how to classify and assess AI systems that serve multiple functions across these categories.

4. AI tools in summative and formative evaluation: What AI tools are most commonly used in both summative and formative evaluation in education? Where would you draw the line between these two types of evaluation? Do you think there are different risk profiles associated with each type of evaluation, and if so, how do they differ?

As the market is evolving quickly, we won't comment on specific products; instead, we'll identify types of assessment features/functions. As noted in our response to Question 1, many of these features are being bundled into existing EdTech platforms and tools.

Summative features/functions: Automated essay grading, standardized test scoring systems, final exam proctoring software, graduation requirement tracking

Formative features/functions: Adaptive learning platforms, real-time feedback tools, AI tutors, learning analytics dashboards, and interim assessment systems.

In education, it's important to note that the distinction between summative and formative assessment (summative evaluation determines final judgments (grades, advancement, credentials). In contrast, formative evaluation guides ongoing learning) is often blurred as continuous assessment models use cumulative formative data for summative decisions.

Risk profiles:

Summative risks: Higher stakes, less opportunity for correction, greater impact on life opportunities, potential for systematic bias in high-stakes decisions

Formative risks: Privacy/data concerns from continuous monitoring, ineffective or incorrect feedback, feedback loops that reinforce biases, and the potential to narrow instruction to what AI can measure

Both carry significant risks. Formative systems may appear "lower risk" but their cumulative data feeds summative decisions and shapes long-term learning trajectories.

5. Interplay with prohibited practices (Article 5): With reference to AI systems listed in Annex III, point 3(d), what AI tools are commonly used to monitor and detect prohibited behaviour during tests and to which extent these tools can be considered as exploiting vulnerabilities or inferring emotions in a way that could cause harm to children?

AI proctoring and monitoring software remain somewhat nascent but are becoming increasingly common.

These tools raise serious concerns under Article 5:

Examples include eye tracking software, facial recognition systems, behavior analysis tools, voice analysis, and emotion recognition during assessments. All of these claim to detect cheating or anomalous behavior.

Exploitation of vulnerabilities (Article 5(1)(a)). These tools exploit students' vulnerable position in test-taking contexts where:

- High stakes create psychological pressure

- False positives cause anxiety that impairs performance

Emotion inference harm (Article 5(1)(f)). Many proctoring tools claim to infer emotions, stress, or deception from physical cues, which:

- Lacks scientific validity

- Causes material harm through false accusations

- Disproportionately impacts students with disabilities, neurodivergent students, and those from different cultural backgrounds

- Creates hostile testing environments

Given this level of concern, these systems should be considered prohibited practices rather than merely high-risk.

6. National or regional guidelines on AI in education: Are you aware of any national or regional official guidelines on the use of AI in schools within the EU? What is your view on the impact of these guidelines, and how might they influence the development and deployment of AI systems in education?

There are emerging guidelines in several EU member states, though implementation varies widely. Some countries have issued general AI ethics frameworks for education, though few have specific, binding regulations governing the use of AI systems in schools. Guidance often focuses on teacher professional development rather than system requirements.

However:

- Guidelines without enforcement mechanisms have limited effect

- Many guidelines don't address the technical requirements needed for compliance with AI Act obligations

- Schools lack resources and expertise to evaluate AI systems against guidelines

There's a gap between high-level principles in existing guidelines and the practical technical requirements of the AI Act. More specific guidance is needed on areas including conformity assessment, risk management, and transparency obligations for educational AI systems.

Any other remarks?

Common Sense Media's AI Risk Assessment Methodology: Our organization has developed a comprehensive framework for evaluating AI systems' safety for youth, which we've applied to major AI platforms including ChatGPT, Gemini, Meta AI, and others. This methodology involves systematic testing across multiple risk categories including: Privacy and data protection practices, content safety and age-appropriateness, potential for manipulation or exploitation, accuracy and reliability of outputs, and transparency and user control features. Our assessments reveal that many widely-used AI systems demonstrate significant safety gaps when used by young people, particularly around mental health support, emotional support, and companionship. This empirical evidence underscores the importance of robust high-risk classification and oversight for educational AI systems.

Chatbots used for mental health support: Our research, conducted in partnership with Stanford Medicine's Brainstorm Lab, shows that AI chatbots for student mental health support are unsafe for teen emotional support despite some improvements in crisis detection. The increasing deployment of tools that provide this kind of support in educational contexts requires urgent regulatory attention.

The educational sector needs clearer guidance on which common AI use cases fall under high-risk classification. Common use cases currently include: Adaptive learning platforms that personalize content, content generation tools used by teachers and students, administrative automation systems, and learning analytics systems. Based on our assessment methodology, we recommend that regulatory guidance explicitly address these common use cases to prevent compliance gaps.

Closing remarks

Do you want to receive updates on this or similar studies?

- ☒ Yes
☐ No

* Please provide us with your email address

afazlullah@commonsense.org

Contact

[Contact Form](#)

