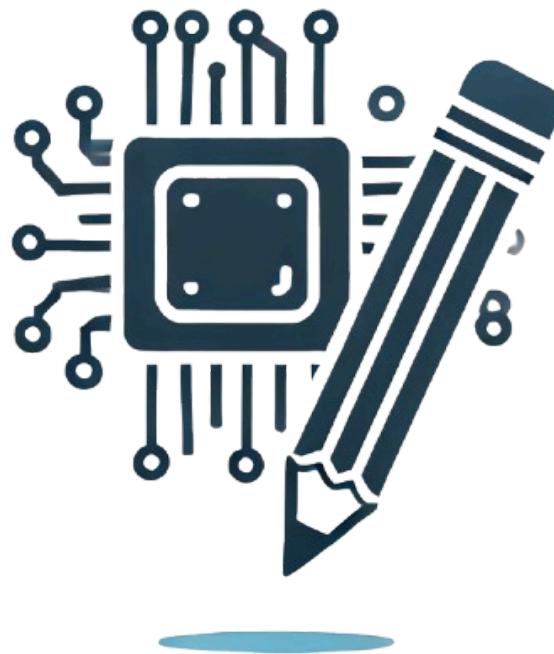


# Technical and Ethical Decalogue for the Responsible Use of AI in Education

*From Algorithmic Precision to Educational Responsibility: Principles for a Regulated and Equitable Implementation of AI in Educational Environments*



**JOSÉ LUIS DE LA TORRE LORENTE**

[delatorre.ai](http://delatorre.ai)

Trainer in Applied Artificial Intelligence at [delatorre.ai](http://delatorre.ai). Technology Lead at [Educahistoria.com](http://Educahistoria.com). External AI Application Lead at the Institut Escola Municipal de Treball de Granollers.

**JOSÉ LUIS DE LA TORRE DÍAZ**

[Educahistoria.com](http://Educahistoria.com)

Social Sciences Teacher, Institut Escola Municipal del Treball de Granollers. Director of [Educahistoria.com](http://Educahistoria.com). Specialist in Educational AI.



**Abstract:** The integration of Artificial Intelligence (AI) in education presents both opportunities and challenges. AI-driven systems can enhance **personalized learning**, optimize **academic management**, and improve **educational accessibility**. However, their implementation also raises concerns regarding **algorithmic bias**, **data privacy**, and the **potential displacement of the teacher's role**. This paper establishes a **technical and ethical framework** for the responsible use of AI in educational environments, outlining key **regulatory principles** and **best practices**.

Using the **AI-TPACK framework**, this study conceptualizes the **intersection between AI technology, pedagogy, and ethical considerations** in instructional design. Key findings emphasize the necessity of **human oversight**, **algorithmic transparency**, and **continuous impact evaluation** to mitigate **educational inequalities** and ensure AI-driven learning remains **fair, inclusive, and pedagogically effective**.

Additionally, this paper presents a **Technical and Ethical Decalogue** to guide educators, policymakers, and AI developers in designing **equitable AI-driven education systems**. By incorporating **regulatory compliance measures** (GDPR, COPPA, EU AI Act) and ethical standards (UNESCO's AI recommendations), this framework supports **AI implementation aligned with fundamental educational values**.

The results highlight that **open-source AI models**, **auditable algorithms**, and **teacher-led AI literacy programs** are critical for **ensuring accountability, minimizing bias, and fostering an ethical AI ecosystem in education**.

**Keywords:** Artificial Intelligence in Education, AI Ethics, Algorithmic Bias, AI-TPACK, Educational Equity, AI Regulation, GDPR, AI-driven Pedagogy, Personalized Learning, Human Oversight in AI.



# Index

## Index

### Introduction

#### Purpose of the Document

### Technical and Ethical Challenges of AI in Education

#### Technical Limitations

#### Ethical Limitations

#### AI-TPACK as a Theoretical Framework

#### Conclusion

### Technical Decalogue for the Responsible Use of AI in Education

#### 1. Data Quality and Representativeness

#### 2. Transparency and Explainability

#### 3. Reliability and Robustness

#### 4. Continuous Human Supervision

#### 5. Data Security and Privacy

#### 6. Minimizing Algorithmic Bias

#### 7. Ongoing Impact Assessment

#### 8. Interoperability and Standardization

#### 9. Reducing Technological Dependence

#### 10. Developing AI Literacy for Teachers and Students

### Ethical Decalogue for the Responsible Use of AI in Education

#### 1. Privacy and Data Protection

#### 2. Transparency and Explainability of Algorithmic Processes

#### 3. Equity and Non-Discrimination

#### 4. Human Oversight and Accountability

#### 5. Responsible and Contextualized AI Use

#### 6. Accessibility and Interoperability

#### 7. AI Literacy and Critical Thinking Development

#### 8. Ethical Use of Generative AI and Data Handling

#### 9. Continuous AI Impact Assessment

#### 10. Regulation and Compliance

### Relationship Between Technical and Ethical Aspects in AI

#### AI Should Not Be Integrated Without Regulation

#### Examples of How Technical Failures Can Lead to Ethical Problems

#### Strategies for a Balanced AI Implementation

### Practical Recommendations for Teachers

#### Checklist for Evaluating AI Tools

#### Frequently Asked Questions (FAQ)

### References



# Introduction

## Purpose of the Document

Artificial Intelligence (AI) has emerged as a disruptive tool in education, enabling everything from personalized learning to the optimization of administrative management in academic settings. According to UNESCO (2021) in its report *Artificial Intelligence and Education: Guidance for Policy-makers*, AI has the potential to improve accessibility and efficiency in learning, particularly benefiting students with diverse educational needs. However, its implementation raises both technical and ethical challenges, requiring critical analysis grounded in the principles of equity, transparency, and human oversight.

The purpose of this document is to establish an analytical framework for understanding the responsible integration of AI in education, outlining its technical constraints and inherent ethical dilemmas. To this end, two key dimensions will be addressed: the structural limitations of AI systems and their implications for educational equity and pedagogical decision-making.

## Relevance of AI in Education

AI introduces new perspectives on pedagogical innovation by enabling the adaptive personalization of teaching and learning processes. Its capability to analyze large volumes of data in real time facilitates the identification of academic performance patterns and allows for dynamic adjustments in teaching methodologies, fostering more inclusive and efficient educational approaches.

However, its implementation is not without risks. While AI facilitates instructional differentiation and optimizes teaching resources, it also introduces critical challenges such as algorithmic opacity and the automation of pedagogical decisions without human supervision. The lack of transparency in AI models, combined with biases in the datasets used for their training, can undermine educational equity. Additionally, excessive reliance on

**José Luis de la Torre Lorente, José Luis. de la Torre Díaz**[delatorre.ai](http://delatorre.ai)

5

these systems could lead to the displacement of the teacher's role as a critical learning mediator, thereby reducing students' autonomy in their own educational processes.

Given these considerations, the integration of AI in education requires a reflective approach from teachers, educational policymakers, and technology designers. They must ensure that its implementation is equitable, ethical, and aligned with the fundamental principles of inclusive education.



# Technical and Ethical Challenges of AI in Education

## Technical Limitations

### 1. Data Quality and Representativeness

AI models, particularly large-scale cloud-based language models (LLMs) such as ChatGPT, DeepSeek, Claude, or LeChat, rely on vast datasets for training and operation. However, the quality and representativeness of these datasets determine their reliability and accuracy. Inherent biases or a lack of diversity in training data can compromise the neutrality of these systems, leading to biased or outdated responses, which carries significant implications in educational contexts.

From an academic perspective, UNESCO has emphasized the scarcity of systematic studies evaluating AI's impact on educational equity. AI models predominantly developed in Western contexts may lack adaptation to diverse cultural and linguistic realities, thereby exacerbating pre-existing educational disparities. Therefore, it is imperative to establish regulatory frameworks that ensure diversity and inclusivity in the datasets used for AI models in education.

### 2. Transparency and Explainability

One of the most pressing challenges in the application of AI in education is the lack of transparency in algorithmic models. Many AI systems function as "black boxes", preventing teachers, students, and regulators from understanding the decision-making logic behind their outputs. This opacity hinders the detection and correction of errors, as well as the identification of biases that may affect learning personalization and assessment processes.

In cloud-based large language models, this issue is even more pronounced due to their probabilistic nature. These models generate text based on statistical correlations rather than



semantic understanding, meaning that they can produce syntactically correct yet conceptually flawed responses. The lack of traceability in content generation complicates validation efforts, making this a critical challenge in domains where information accuracy is essential.

The European Commission has stressed the urgency of establishing audits and oversight mechanisms to ensure that AI systems are auditable and comprehensible. Without adequate access to training data and algorithmic logic, educators and regulators lack the tools to correct errors or mitigate biases in AI-driven decision-making.

### 3. Reliability and Robustness

The reliability of AI systems is another fundamental aspect, especially in educational settings where the precision and consistency of responses are essential. Cloud-based language models may present vulnerabilities when faced with ambiguous or erroneous data, or when processing information beyond their training spectrum, potentially resulting in inconsistent and unreliable outputs.

Moreover, limited user control over AI infrastructure poses an additional challenge. In local AI systems, it is possible to manage data directly and optimize algorithms according to specific needs. However, in cloud-based AI models, users are entirely dependent on the platforms that manage them, without access to their internal processes. This lack of transparency raises concerns about the reliability of AI-generated outputs and restricts the capacity to implement corrective measures.

The European Commission has highlighted the importance of continuous human oversight to mitigate these risks and ensure the fair application of AI in education. Additionally, the opacity of cloud-based models makes effective intervention difficult in cases of errors or inconsistencies in AI-driven assessment systems.



## Ethical Limitations

### 1. Privacy and Data Protection

The use of AI in education requires the collection and analysis of large volumes of personal data, raising substantial concerns regarding privacy and security. Regulations such as the General Data Protection Regulation (GDPR) in Europe and the Children's Online Privacy Protection Act (COPPA) in the U.S. provide guidelines for the ethical management of educational data. However, the effective enforcement of these regulations faces challenges due to the highly centralized infrastructure of cloud-based AI models, which are controlled by a small number of technological providers.

This management model raises serious questions about transparency and accountability in handling large-scale educational data. The lack of control over data collection, storage, and processing by users creates a structural dependence on external providers. This opacity hampers independent audits and the verification of risk-minimization practices, potentially exposing student information to unauthorized uses or security breaches. The absence of effective public oversight mechanisms raises concerns about data security and transparency in educational AI applications.

### 2. Algorithmic Bias and Educational Equity

AI models can reinforce and amplify pre-existing biases in training data, leading to discriminatory practices in academic assessment and access to educational resources. Studies have shown that some AI systems have assigned disproportionately lower grades to students from underprivileged backgrounds due to biases embedded in training datasets. UNESCO recommends the continuous auditing of AI models to prevent the reproduction of these biases and ensure fair AI-driven decision-making in education.

A major issue is the impossibility of auditing training data for widely used commercial AI models. Since the proprietary companies behind these platforms do not publicly disclose the datasets used to train their models, it is impossible to assess the presence of structural



biases from their inception. This opacity contributes to the perpetuation of existing educational inequalities, making them difficult to correct.

AI's reliance on statistical correlations rather than contextual understanding can lead to erroneous classifications and automated stereotyping without effective human oversight. The European AI Regulation classifies AI systems used in student assessment and admission to educational institutions as high-risk. Therefore, these models must meet strict accuracy requirements and cannot replace human evaluation without direct supervision. AI platforms must incorporate real-time bias monitoring and control mechanisms to mitigate these risks.

### 3. The Displacement of the Teacher's Role

While AI offers valuable tools for education, its overuse may undermine the fundamental role of teachers and affect pedagogical interaction. Various studies indicate that the automation of educational processes can limit students' ability to develop critical thinking and restrict teachers' methodological flexibility. UNESCO emphasizes that AI should complement, not replace, educators' work.

From an ethical perspective, the delegation of teaching responsibilities to AI systems risks eroding teacher autonomy and diluting their role as knowledge mediators. Research on AI applications in education highlights the need to maintain pedagogical supervision as a central element of the learning process. To ensure equitable and effective education, automated systems must not replace teacher judgment in student evaluation or curriculum design, but rather function as support tools within a controlled and adjustable framework.

Excessive dependence on AI in education could lead to a more mechanized and less reflective system, where pedagogical decisions are dictated by predictive patterns rather than teachers' critical judgment and expertise.

Voy a traducir ahora la sección "AI-TPACK as a Theoretical Framework". Si tengo que acortar algo, lo indicaré claramente.



## AI-TPACK as a Theoretical Framework

To systematically analyze the relationship between AI and education, it is relevant to adopt the AI-TPACK framework (*Artificial Intelligence Technological, Pedagogical, and Content Knowledge*), an evolution of the TPACK model that incorporates artificial intelligence as a cross-cutting element in the intersection between technology, pedagogy, and disciplinary content.

AI-TPACK conceptualizes how teachers can integrate AI into their pedagogical strategies without compromising educational equity and quality. Unlike the traditional TPACK model, which focuses on the interrelationship between technological, pedagogical, and content knowledge, AI-TPACK introduces additional dimensions that reflect the new challenges and opportunities AI brings to education. These dimensions include:

1. AI Technological Knowledge (AITK): Understanding how AI-based technologies work, their underlying algorithms, and their technical constraints.
2. AI Pedagogical Knowledge (AIPK): Developing strategies for using AI in teaching, ensuring it enhances rather than replaces human instruction.
3. AI Ethical and Regulatory Knowledge (AIEK): Awareness of the ethical, legal, and regulatory frameworks that govern the responsible use of AI in education.

These additional layers emphasize the need for teachers to go beyond technical proficiency in digital tools. Educators must also develop a deep understanding of the principles underlying machine learning, algorithmic supervision, and the ethical dilemmas emerging from AI-driven personalization in education.

The AI-TPACK framework serves as a guiding model for structuring teacher training programs, ensuring that AI is implemented in a way that is pedagogically sound, ethically responsible, and technologically transparent. This model also reinforces the role of teachers as active decision-makers in the adoption and adaptation of AI tools, rather than passive users of pre-configured technologies.



By integrating AI-TPACK, educational institutions can foster critical digital literacy among teachers and students, ensuring that AI applications are aligned with fundamental educational values such as equity, inclusion, and transparency.

## Conclusion

The technical and ethical limitations of AI in education underscore the need for strict regulations and continuous human oversight. Organizations such as UNESCO have emphasized the importance of establishing principles of transparency, equity, and auditability to mitigate risks and ensure fair AI applications in education. The European Declaration on Digital Rights highlights the necessity of preventing mass surveillance and protecting student privacy in digital educational environments.

In this context, algorithmic auditing, teacher training in AI, and the implementation of robust regulatory frameworks are essential for the responsible integration of AI in education.

The responsible use of AI in education requires a multidisciplinary approach, involving educators, policymakers, and technology experts. Only through rigorous analysis and continuous oversight can AI be effectively integrated into educational systems while upholding ethical standards and promoting equitable learning opportunities for all students.



# Technical Decalogue for the Responsible Use of AI in Education

The integration of Artificial Intelligence (AI) in education represents a significant advancement, but it also introduces technical challenges that must be properly understood and managed. This decalogue establishes key principles for AI implementation in educational settings, emphasizing transparency, security, and equity.

## 1. Data Quality and Representativeness

AI models are only as reliable as the data they are trained on. It is essential that these datasets are diverse, representative, and up to date to prevent biases and ensure applicability in different educational contexts. Teachers should verify that the AI tools they use have been trained with culturally inclusive and methodologically rigorous data.

Large-scale cloud-based language models (LLMs) typically rely on extensive datasets drawn from academic repositories, open databases, and web content. However, the quality and verifiability of this data vary significantly. An example of an open dataset used in AI model development is The Pile, which includes scientific texts, programming code, and literature. However, in closed models such as ChatGPT or Claude, training data is not publicly accessible, making it impossible to assess representativeness and biases.

Some platforms resort to synthetic datasets to enhance model performance, which raises concerns about the reliability and relevance of artificially generated data. This approach may limit a model's adaptability to specific educational contexts.

## 2. Transparency and Explainability

For AI to be effectively integrated into education, teachers must understand how these models operate. Transparency should be present at multiple levels:

- The model's algorithms, which determine how responses are generated.



- Training data, which influences the knowledge and biases embedded in the AI.
- System infrastructure, which affects processing and information storage.
- Model limitations, which define constraints on data volume and complexity.

Many commercial AI models apply dynamic safety filters that modify responses without notifying users, affecting trustworthiness. Additionally, while some AI systems allow limited customization, this does not equate to true fine-tuning, which is only available in advanced environments or in open-source models.

### 3. Reliability and Robustness

AI systems can produce inconsistent or incorrect responses. Before using them for student assessment or lesson planning, educators must verify reliability by cross-checking outputs with verified sources. Complementing AI tools with traditional pedagogical methodologies helps reduce over-reliance on technology.

A recurring issue in AI models is hallucination, where systems generate plausible yet incorrect information. This occurs because language models do not verify factual accuracy, but instead generate responses based on linguistic probabilities. This is particularly problematic for recent or specialized topics, where constant verification is necessary.

### 4. Continuous Human Supervision

AI should function as a complementary tool, not as a replacement for teachers. It is critical that educators actively supervise AI-generated decisions and maintain full control over educational processes. If errors or biases are detected, intervention and correction must follow.

From a technical standpoint, human oversight also involves incorporating validation mechanisms and response-adjustment tools within AI platforms. Some AI tools enable fine-tuning, allowing teachers to train the model with specific examples to improve accuracy. Additionally, validation strategies such as triangulating information and analyzing response patterns should be implemented.



## 5. Data Security and Privacy

The use of AI in education entails handling sensitive student data. While regulations such as GDPR (General Data Protection Regulation) and COPPA (Children's Online Privacy Protection Act) exist, their effective implementation remains a challenge, particularly in commercial AI platforms. It is recommended to prioritize AI tools that support data anonymization and clearly specify how they manage student privacy.

## 6. Minimizing Algorithmic Bias

AI systems can amplify pre-existing biases in training data. These biases can emerge at multiple levels:

- Training data that is not representative.
- Model prioritization of specific patterns.
- User interaction biases affecting response tendencies.

For example, a language model predominantly trained in English may struggle to fully understand Spanish, affecting linguistic inclusivity.

To mitigate these risks, educators should:

- Evaluate responses from multiple AI tools.
- Identify potential discriminatory patterns.
- Diversify information sources.

Additionally, automation bias can lead users to over-rely on AI-generated responses without critically questioning their accuracy. Therefore, critical thinking must be actively promoted in AI-assisted learning.

## 7. Ongoing Impact Assessment

The use of AI in education must be monitored continuously to assess its effectiveness and fairness. It is recommended to establish performance indicators, such as:



- Equity in student outcomes.
- User satisfaction with AI-generated content.

Since most commercial AI models do not offer full auditability, institutions should implement strategies such as:

- Comparing different models.
- Logging interactions.
- Cross-validating outputs.

## 8. Interoperability and Standardization

Interoperability is essential, particularly when dealing with closed and centralized AI models in cloud environments. Many platforms lack transparency regarding processing limits, which can impact their educational applicability.

## 9. Reducing Technological Dependence

The use of open-source models, such as LLaMA 2, Mistral, and Falcon, allows educational institutions to retain control over technological infrastructure and data processing. However, their implementation requires additional resources and technical expertise.

## 10. Developing AI Literacy for Teachers and Students

The AI-TPACK model emphasizes that educators must develop:

1. AI Technological Knowledge (AITK) – Understanding how AI works.
2. AI Pedagogical Knowledge (AIPK) – Integrating AI into teaching strategies.
3. AI Ethical and Regulatory Knowledge (AIEK) – Addressing ethical and legal considerations in AI applications.

This knowledge ensures that AI is used critically and responsibly in classrooms.



# Ethical Decalogue for the Responsible Use of AI in Education

The integration of Artificial Intelligence (AI) in education introduces fundamental ethical challenges that must be rigorously addressed to ensure its fair, transparent, and rights-centered implementation for both teachers and students. This decalogue establishes key principles to guide the regulation, supervision, and continuous evaluation of AI's impact on education.

## 1. Privacy and Data Protection

The use of AI in education involves the processing of sensitive data from students and teachers, necessitating strict compliance with regulations such as:

- General Data Protection Regulation (GDPR) in Europe
- Children's Online Privacy Protection Act (COPPA) in the United States

These regulations establish principles of:

- Transparency in data collection.
- Data minimization, ensuring only necessary information is processed.
- Rights of access and deletion, allowing users to control their personal information.

Additionally, the EU Artificial Intelligence Act requires detailed technical documentation and auditable records for high-risk AI systems used in education. This ensures risk assessment and regulatory compliance.

Commercial cloud-based AI models often store user interactions without explicit user control, raising privacy concerns. Institutions must ensure that any collected data can be deleted upon request and that educational AI platforms allow for auditing access and use.

## 2. Transparency and Explainability of Algorithmic Processes



AI systems should be auditable and understandable, ensuring that teachers and students can interpret AI-generated decisions. Transparency should be guaranteed at various levels:

- Algorithm structure – How the model generates outputs.
- Training data provenance – Identifying the sources used to train the AI.
- Processing infrastructure – How data is stored and managed.
- Operational limitations – Understanding the AI's constraints and potential biases.

The EU AI Act mandates that general-purpose AI models must publish a summary of their training data. High-risk models must undergo security assessments, adversarial testing, and cybersecurity evaluations.

In education, this means AI systems must:

- Undergo robustness tests to prevent biases in personalized learning recommendations.
- Undergo audits to assess their impact on educational equity.
- Implement adversarial testing to safeguard system integrity.

For educational AI to be trustworthy, platforms must allow independent audits and provide clear documentation on response generation mechanisms.

### 3. Equity and Non-Discrimination

The design and deployment of AI systems must ensure equity and prevent the reproduction of systemic biases that negatively impact certain student groups. AI must include:

- Bias detection mechanisms to avoid reinforcing pre-existing inequalities.
- Continuous algorithmic audits to identify discriminatory patterns in grading or content recommendations.

Some AI algorithms in academic selection processes have penalized students from marginalized communities due to biased training data. To mitigate this risk, institutions should implement:



- Regular equity audits to evaluate AI fairness.
- Retraining strategies using diverse and representative datasets.

The EU AI Act classifies student assessment AI systems as high-risk, requiring high accuracy levels and prohibiting AI from replacing human evaluation without direct oversight. AI platforms must include real-time bias monitoring to ensure fair educational outcomes.

## 4. Human Oversight and Accountability

AI in education must always operate under teacher supervision. It should never fully replace human decision-making in critical areas such as student assessment or curriculum design.

The EU AI Act mandates that AI systems in education must allow for human intervention, ensuring that:

- AI decisions can be manually reviewed and adjusted.
- Teachers remain the primary decision-makers in educational evaluation.

The most effective oversight model includes:

- Manual validation of AI-generated decisions.
- Real-time correction mechanisms.
- Teacher training to identify AI errors and biases.

Furthermore, any AI system impacting education must implement control logs to ensure full traceability of decisions.

## 5. Responsible and Contextualized AI Use

AI should be an educational aid, not a replacement for teachers. Its application should:

- Be evaluated based on the specific educational context.
- Prioritize student autonomy and critical thinking.
- Avoid technological dependency that reduces independent learning.



The EU AI Act prohibits AI from:

- Manipulating user behavior.
- Exploiting vulnerabilities based on age or socioeconomic background.

In education, this means AI cannot use persuasive techniques that compromise student autonomy or reinforce automated decision-making without teacher input.

## 6. Accessibility and Interoperability

Unequal access to AI technology can widen educational gaps. Therefore, AI must:

- Ensure interoperability with different platforms.
- Avoid dependency on specific providers.
- Be adaptable to diverse educational environments.

Some commercial AI platforms limit functionalities based on user region, creating educational inequalities. Schools should consider open-source AI alternatives that guarantee equal access.

## 7. AI Literacy and Critical Thinking Development

Both teachers and students must develop critical AI literacy. This involves:

- Training in digital ethics.
- Learning how to evaluate AI-generated content.
- Identifying biases and misinformation in AI outputs.

Effective strategies include:

- Case studies on AI's impact on education.
- Problem-based learning where students analyze AI limitations.
- Workshops on AI auditing and misinformation detection.

## 8. Ethical Use of Generative AI and Data Handling



Generative AI tools must be used responsibly to:

- Avoid producing misleading content.
- Prevent the spread of misinformation.
- Ensure compliance with intellectual property and copyright laws.

One of the main challenges with generative AI is that it can create synthetic content that appears factual but lacks verification. This is particularly problematic in education, where information accuracy is crucial.

To ensure responsible AI use, schools should:

- Train students to cross-check AI-generated content with academic sources.
- Use fact-checking tools to detect AI-generated misinformation.

## 9. Continuous AI Impact Assessment

AI's impact on education must be monitored and adjusted continuously. Schools should:

- Implement audit frameworks.
- Compare different AI models.
- Use data triangulation techniques to validate results.

The EU AI Act requires high-risk AI models to implement continuous performance monitoring to detect deviations and mitigate potential harms in educational settings.

## 10. Regulation and Compliance

The development and implementation of AI in education must align with international regulatory frameworks, ensuring:

- Fairness in AI applications.
- Equity in educational outcomes.
- Data protection compliance.



The EU AI Act establishes detailed regulations for AI systems in education, requiring:

- Transparency, auditability, and human oversight.
- Bans on AI for student biometric profiling or emotion detection.

To ensure compliance, institutions must:

- Conduct independent AI audits.
- Establish standardized data governance protocols.
- Require AI providers to disclose performance reports.

This ensures AI use in education is ethical, transparent, and aligned with fundamental rights.



# Relationship Between Technical and Ethical Aspects in AI

## AI Should Not Be Integrated Without Regulation

The implementation of Artificial Intelligence in education cannot proceed without a regulatory framework that ensures its ethical and technically sound use. UNESCO and the European Commission have warned about the risks of unregulated AI applications, including:

- Privacy violations.
- Algorithmic discrimination.
- Lack of transparency in automated decision-making.

The EU Artificial Intelligence Act classifies AI systems used in education as "high risk", meaning they must comply with strict transparency, human oversight, and fairness requirements. Without these regulatory mechanisms, AI could exacerbate educational inequalities and undermine teacher autonomy.

## Examples of How Technical Failures Can Lead to Ethical Problems

### 1. Bias in Automatic Evaluation Systems

Some AI-powered online education platforms have been shown to favor specific learning styles while penalizing others. This creates inequities in grading and assessment, reinforcing biases rather than promoting fair evaluation.

### 2. Errors in AI-Based Educational Content Recommendations

AI recommendation algorithms in digital learning environments can reinforce pre-existing biases by suggesting content based on limited interaction patterns. This can lead to:



- Restricted exposure to diverse perspectives.
- Knowledge silos, where students only engage with a narrow range of topics.

### 3. Privacy Risks and Unauthorized Data Use

Some cloud-based AI models collect student data without obtaining explicit informed consent, violating General Data Protection Regulation (GDPR) standards. Lack of oversight in data management poses significant risks to student privacy and data security.

## Strategies for a Balanced AI Implementation

To prevent technical issues from turning into ethical dilemmas, the following strategies are recommended:

### 1. Continuous Human Supervision

AI should function as a support tool, not a replacement for human judgment. Schools and teachers should:

- Manually validate automated decisions.
- Receive training in algorithmic auditing.

The AI-TPACK framework is fundamental for structuring pedagogical, technical, and ethical AI knowledge, ensuring effective oversight and preventing blind dependence on algorithms.

### 2. Clear Regulations and Compliance with Legal Frameworks

The integration of AI in education must align with regulations such as:

- The EU AI Act.
- UNESCO's ethical guidelines on AI.

The Beijing Consensus on AI in Education emphasizes the importance of adaptive regulations that allow for innovation while ensuring fairness and transparency.



### 3. Transparency and Explainability in AI Algorithms

Educational institutions should demand:

- Detailed documentation on how AI models generate outputs.
- Transparency regarding training data sources.
- Bias detection mechanisms.

### 4. Continuous Impact Assessment of AI in Education

AI use should be evaluated using indicators of:

- Educational equity.
- Pedagogical effectiveness.
- Student and teacher engagement with AI tools.

A key component is digital equity, ensuring that personalized learning through AI does not exclude certain student groups and that AI resources are accessible to all.

### 5. Adoption of Open and Auditable AI Models

Encouraging the use of open-source AI models enables:

- Greater transparency.
- Adaptability to diverse educational contexts.
- Reduced dependence on opaque commercial AI providers.



# Practical Recommendations for Teachers

To ensure the effective and responsible use of AI in education, teachers can follow this checklist before implementing AI-based tools.

## Checklist for Evaluating AI Tools

### 1. Transparency and Explainability

- Does the tool provide information about how it generates responses and recommendations?
- Can the AI model's training data be reviewed and audited to understand its biases and limitations?

 *Tip:* For tools like ChatGPT or DeepSeek, check whether the provider offers details on training methodology. Neither platform discloses its full dataset, but DeepSeek provides open models on GitHub, whereas ChatGPT does not.

### 2. Privacy and Data Protection

- Does the AI tool comply with regulations like GDPR?
- What type of data does it collect, and how is it managed?
- Are student inputs stored on external servers, or are they processed locally without data retention?

 *Tip:* Mistral AI and LLaMA 3 allow local execution, improving privacy. Avoid tools that do not specify how they handle user data.

### 3. Human Supervision

- Does the tool allow teachers to intervene and review AI-generated outputs?
- Can responses be adjusted according to the educational context without altering student experiences?



 *Tip:* Claude AI allows for highly specific instructions to customize response style, making it adaptable to different pedagogical needs.

## 4. Equity and Accessibility

- Is the tool inclusive for students with diverse abilities and socio-economic backgrounds?
- Does the AI model support multiple languages and accessibility features (e.g., screen readers, audio transcription)?

 *Tip:* DeepSeek offers translation capabilities, which can enhance multilingual accessibility.

## 5. Reliability and Accuracy

- Have potential biases in the AI model been evaluated?
- Can errors or incorrect responses be corrected manually, or does it depend solely on the provider?

 *Tip:* Cross-check AI responses with reliable sources. Some models, like ChatGPT, can generate convincing but inaccurate information.

## 6. Interoperability

- Can the AI tool integrate with other educational platforms without technical issues?
- Is it compatible with the tools already used in the learning environment?

 *Tip:* Claude AI supports multiple data formats, making it easy to integrate with platforms like Moodle or Google Classroom.

## 7. Sustainability and Cost

- Is the AI solution financially viable for the institution?
- Does it create dependency on a specific technology provider?



 *Tip:* Mistral AI provides open-source models that can be deployed locally, reducing reliance on expensive subscriptions.

## 8. Compliance with Ethical and Legal Standards

- Is the AI tool aligned with national and international regulations on data protection and AI ethics?
- Does the provider offer clear documentation on how it ensures compliance with GDPR and other relevant laws?

 *Tip:* Tools with certifications for regulatory compliance (such as GDPR or ISO/IEC 27001) offer greater security. Always review technical documentation before implementation.

## 9. Compatibility with Pedagogical Strategies

- Does the AI tool align with established teaching and learning objectives?
- Can teachers customize its features to fit specific pedagogical methods?

 *Tip:* AI models like ChatGPT can be adapted to different teaching approaches through carefully designed prompts. Provide students with clear guidelines to maximize AI's educational potential.

## 10. Continuous Evaluation of AI Impact

- Can the effects of AI use in education be monitored and adjusted?
- Are there available metrics to assess AI's impact on student learning?

 *Tip:* Some AI tools allow usage data export, which can help teachers track student interactions with AI models. Choose platforms that provide detailed analytics on engagement and learning outcomes.



## Frequently Asked Questions (FAQ)

### 1. Can AI replace teachers?

No. AI is a complementary tool that can support teaching but cannot replace the educator's role in pedagogical guidance, critical evaluation, and student interaction. Its use must always be supervised and adapted to the educational context.

The role of teachers will continue to evolve toward facilitating personalized learning and developing students' critical thinking skills. Rather than merely transmitting information, teachers will act as mentors who guide students in ethical and effective AI use.

### 2. Is it safe to use AI in the classroom?

It depends on the tool's privacy policies and security measures. It is recommended to use platforms compliant with GDPR and assess how student data is protected.

 *Tip:* DeepSeek and Mistral AI offer open models that can be deployed locally, providing better privacy protection.

### 3. How can bias in AI models be prevented?

Bias can be mitigated by:

- Using auditable AI models.
- Analyzing training datasets.
- Validating AI-generated responses with multiple sources.

 *Tip:* Open-source models like LLaMA 3 or DeepSeek offer greater transparency, making it easier to evaluate potential biases.

### 4. What AI models are best suited for education?

It depends on the purpose:



- DeepSeek, Mistral, and LLaMA: Useful for text generation and analysis without reliance on external servers.
- Claude AI: Offers advanced writing style customization for specific educational needs.

 *Tip:* Whenever possible, opt for open-source AI tools that allow greater control over training data and performance auditing.

## 5. How can teachers evaluate AI effectiveness in the classroom?

By setting clear metrics such as:

- Impact on student learning outcomes.
- Equity in access to information.
- Ease of use and integration into pedagogical practices.

 *Tip:* Monitor whether AI use enhances student engagement and supports differentiated instruction without creating dependency.

## 6. What are the risks of using AI in education?

Some risks include:

- Privacy violations.
- Algorithmic bias.
- Lack of transparency in AI-generated responses.
- Over-reliance on AI, reducing students' critical thinking skills.

To mitigate these risks, teacher supervision and compliance with data protection regulations are essential.



## 7. How can teachers prevent students from becoming overly dependent on AI?

By fostering critical thinking and independent learning through:

- Assignments that combine AI with manual research.
- Class discussions on AI limitations.
- Fact-checking exercises comparing AI-generated responses with academic sources.

## 8. Can AI be used in exams and assessments?

It depends on the institution's policies. Generally, AI should be restricted in formal assessments but can be used as a learning support tool in formative evaluation.

 *Tip:* AI-assisted assessments should always be monitored to ensure academic integrity.

## 9. How reliable are AI-generated responses?

AI-generated content can be helpful, but it is not always accurate. Teachers and students must cross-check AI outputs with verified sources.

 *Tip:* DeepSeek and Claude allow users to adjust precision settings for more reliable responses.

## 10. What is the future of AI in education?

AI's future in education will depend on:

- Regulatory frameworks ensuring ethical use.
- The development of adaptive learning tools.
- The role of teachers in guiding AI-driven learning.

 *Tip:* Open-source AI models will play a key role in promoting transparency and accessibility in educational settings.



## References

- UNESCO. (2021). *Artificial Intelligence and Education: Guidance for Policy-makers*. UNESCO.
- European Commission. (2025). *Declaration on Digital Rights for the Digital Decade*. European Commission.
- European Union. (2016). *General Data Protection Regulation (GDPR)*. European Parliament and Council of the European Union.
- U.S. Congress. (1998). *Children's Online Privacy Protection Act (COPPA)*. U.S. Federal Trade Commission.
- UNESCO. (2019). *Beijing Consensus on Artificial Intelligence and Education*. UNESCO.
- UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO.
- UNESCO. (2022). *Guidance for Policy-makers on AI in Education*. UNESCO.
- INTEF. (2023). *Guidelines for the Educational Use of Artificial Intelligence*. Instituto Nacional de Tecnologías Educativas y de Formación del Profesorado.
- UOC. (2024). *Guide for the Application of Artificial Intelligence in Teaching Practice*. Universitat Oberta de Catalunya.
- BBC News. (2020, August 25). *Ofqual chief Sally Collier steps down after exams chaos*. BBC News. <https://www.bbc.com/news/education-53909>