



AI@MEDIATORS

**Artificial intelligence and digital educational tools
for the prevention of cyber-bullying**



**Co-funded by
the European Union**



Universidad de Alicante



ERUDITO LICÉJUS





Contents

1	3
1.1.	3
1.2.	3
2	4
2.1.	4
2.2.	5
2.3.	6
3	7
3.1.	7
3.2.	7
3.3.	8
4	1
5	1
6	2

1 Evaluation Schema for AI in Education, task presentation and methodology

1.1. Introduction and general objective

In the contemporary digital landscape, characterized by ubiquitous technological integration, **the conscientious deployment of artificial intelligence (AI) within the educational sector emerges as a pivotal concern** in terms of understanding of the opportunities, but also with a required attention on the related risks.

The AI@Mediator project, aiming to explore the opportunity of tackling cyberbullying issues through the implementation of an AI-based chatbot, **stands in the middle of this complex matter**. As members of AI@Mediators Consortium we are called not only to contribute to the debate, but also to provide useful tools and methodologies which can support others who have started later on this reflection and path on the implementation of AI in the field of education, in particular with the vulnerable demographic of minors.

Within the framework of this ambitious initiative, the development of a meticulous evaluation schema for AI in education assumes particular significance. **This document represents our perspective, our element of guidance** to help the readers understanding if the AI tools or practices they are adopting, or they are considering to adopt, have on one side the minimum requirement to be implemented with success, on the other side, if they are compelling with the higher ethical standards in term of privacy and EU values.

1.2. Specific objectives

The evaluation schema aims to provide a pivotal reference, to facilitate the informed assessment and ethical integration of AI technologies within educational contexts. By delineating clear parameters for ethical AI utilization, the schema feeds the project's overarching aim of cultivating a secure online environment and fostering digital resilience among diverse target groups. Moreover, the schema emerges as a potent instrument in advancing the project's mission to empower educators in combating cyberbullying effectively. **Through its provision of comprehensive guidelines and assessment frameworks, the schema equips educational stakeholders with the requisite tools to proactively address instances of cyberbullying and engender a culture of inclusivity and digital citizenship within educational ecosystems.**

- **Suggest an evaluation strategy to discern the theoretical dimension of AI technologies within educational realms**, with a pronounced emphasis on data privacy, ethical considerations, and alignment with European Union values and policies. This approach is strategically positioned to furnish a framework for the discernment of responsible AI deployment, thereby dovetailing seamlessly with the overarching objectives of the AI@Mediator project. Indeed, the overall relevance of this document to the project's success should not be underestimated.
- **Establishing a set of evaluation criteria to measure the impact dimension of the deployed AI tools and methodologies, according to the three criteria of capability, utility and adoption.** This approach aims to contribute to understanding if the deployed approach is the most suitable one in terms of opportunity cost, requirements, target group needs, framework and social/cultural environment. Indeed, a promising tool or methodology can result in a failure in implementation, if the choice is not done taking in consideration the specific requirements and needs of the target group it is going to target.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Project No: 2023-1-ES01-KA220-SCH-000160237

2 Evaluating the ethical dimension

This comprehensive analysis delves into the risks and opportunities associated with three dimensions :data privacy, ethics and EU recommendations. This activity aims to underscore their pivotal role in the successful implementation of AI initiatives, particularly in the realm of cyberbullying prevention. By examining the interplay between data privacy, ethics, and EU values, educational institutions can effectively point out risks, deploy mitigation measures, and foster a culture of responsible AI use. This exploration serves as a foundational framework for guiding strategic decision-making and ensuring that AI@Mediators selected AI best practices aligned with ethical principles, regulatory requirements, and societal values within the educational ecosystem.

In this sense, it is important to understand as any ethical evaluation may be affected by personal or cultural bias, which can result in partially inadequate conclusions. So, any ethical evaluation needs to be well-detailed, framed and commented to be meaningful, and any analysis needs to be updated and refreshed periodically in accordance with the evolving situation and needs.

Here below it is possible to find the suggested evaluation framework.

2.1. Data Privacy

- **Major risks:** One of the primary risks associated with data privacy and AI is the potential for unauthorized access or unaware sharing or leaks of sensitive information, leading to breaches of confidentiality and even personal exposure of minors. Inadequate data protection measures and a lack of general data-wisdom could result in alarming situations, compromising the privacy rights of students and educators and even affecting their daily life. Moreover, mishandling of data could violate regulatory requirements such as the General Data Protection Regulation (GDPR) or other related national laws, exposing educational institutions to legal and financial repercussions.
- **Mitigation measures:**
 - Implement robust data protection measures: Educational institutions should establish stringent protocols for data collection, storage, and processing to safeguard student privacy. This includes encryption, access controls, and regular security audits to prevent unauthorized access and breaches.
 - Obtain informed consent: Prior to collecting any student data, institutions should obtain explicit consent from students or their guardians, outlining the purpose and scope of data usage. Transparency regarding data practices fosters trust and empowers individuals to make informed decisions about their privacy.
 - Minimize data retention: To mitigate the risk of data exposure, institutions should adopt a policy of minimal data retention, only retaining data for as long as necessary to fulfill educational objectives. Regular data purges and deletion protocols should be implemented to remove outdated or unnecessary information.
- **Suggested investigation questions:**
 - How is student data collected, stored, and processed within the AI-based educational tool or practice?
 - Are adequate security measures in place to protect student data from unauthorized access or breaches?
 - Is explicit consent obtained from students or their guardians before collecting and using their data?
 - What procedures are in place for data retention and deletion to minimize the risk of data exposure?

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Project No: 2023-1-ES01-KA220-SCH-000160237



- How does the AI tool or practice ensure data anonymization and privacy preservation throughout its lifecycle?

2.2. Ethics

- **Major risks:** Ethically, the use of AI in education raises significant concerns regarding fairness, transparency, and accountability. AI algorithms may exhibit biases or discriminatory outcomes, perpetuating inequalities and reinforcing existing power dynamics (very sensitive topics if referred to the bullying dynamics). Additionally, the opacity of AI decision-making processes may erode trust and undermine human agency, particularly in sensitive areas such as assessment and student support. Ethical considerations also extend to issues of consent, autonomy, and the potential for AI to infringe upon individual rights and freedoms. Mitigating these risks necessitates the adoption of ethical guidelines and standards that prioritize fairness, transparency, and human-centered design principles in AI development and deployment.
- **Mitigation measures:**
 - Conduct bias assessments: Prior to deploying AI systems, educational institutions should conduct thorough assessments to identify and mitigate biases in algorithms and data sources. This may involve auditing training data, testing for disparate impact, and implementing bias-mitigation strategies to ensure fair and equitable outcomes.
 - Enhance transparency: Institutions should prioritize transparency in AI decision-making processes, providing clear explanations of how algorithms operate and the factors influencing their outputs. Transparent AI systems empower stakeholders to understand and challenge decisions, promoting accountability and trust.
 - Establish ethical review boards: Educational institutions should establish interdisciplinary review boards tasked with evaluating the ethical implications of AI initiatives. These boards can provide guidance on ethical dilemmas, assess proposed AI applications, and ensure compliance with ethical standards and regulations.
- **Suggested investigation questions:**
 - How does the AI-based educational tool or practice address potential biases in its algorithms and data sources?
 - Is there transparency regarding the decision-making processes of the AI tool, including factors influencing its outputs?
 - Are there mechanisms in place to ensure fairness, accountability, and human oversight in AI-driven decision-making?
 - How does the AI tool or practice promote ethical principles such as autonomy, respect for individual rights, and non-discrimination?
 - Has the AI tool undergone ethical review or assessment to evaluate its compliance with ethical standards and guidelines?

2.3. EU Recommendations

- **Major risks:** From the perspective of EU recommendations, the use of AI in education must align with regulatory frameworks and policy guidelines established by the European Union. EU recommendations emphasize the importance of upholding fundamental rights and values, including data protection, privacy, and non-discrimination, but also more specific and technical aspects related to the anticipated AI-Act, which is expected to enter in force in 2024. Overall, educational institutions are expected to comply with regulations such as the General Data Protection Regulation (GDPR) and adhere to ethical principles outlined in documents like the Ethics guidelines for trustworthy AI. It is crucial to notice as this effort should be considered continuous, since the EU framework for AI regulation is quickly evolving. Failure to meet these standards may result in legal non-compliance, reputational damage, and loss of trust among stakeholders. Therefore, educational stakeholders must prioritize EU recommendations and integrate them into their AI governance frameworks to ensure responsible and ethical AI utilization in education.
- **Mitigation measures:**
 - Adhere to GDPR compliance: Educational institutions must ensure compliance with the General Data Protection Regulation (GDPR) and other relevant data protection laws. This includes implementing measures such as data anonymization, privacy impact assessments, and procedures for handling data breaches in accordance with GDPR requirements.
 - Follow European AI Ethics Guidelines: Institutions should align their AI practices with the European AI Ethics Guidelines, which outline ethical principles and recommendations for AI development and deployment. By adhering to these guidelines, institutions demonstrate their commitment to ethical AI governance and alignment with EU values.
 - Engage in ongoing training and awareness: Educational stakeholders should invest in training programs and awareness initiatives to ensure that staff and students are informed about their rights and responsibilities regarding AI usage. By fostering a culture of ethical awareness and compliance, institutions can uphold EU recommendations and promote responsible AI utilization in education.
- **Suggested investigation questions:**
 - Does the AI-based educational tool or practice comply with relevant data protection regulations such as the GDPR?
 - Is there alignment with the European AI Ethics Guidelines in the development and deployment of the AI tool or practice?
 - Are measures in place to uphold fundamental rights and values outlined in EU recommendations, including data protection, privacy, and non-discrimination?
 - How does the AI tool or practice demonstrate accountability and transparency in accordance with EU regulatory requirements?
 - Has the institution engaged in ongoing training and awareness initiatives to ensure compliance with EU recommendations and promote responsible AI utilization?

3 Evaluating the impact dimension

In the expanding field of application of AI based solutions, the integration of artificial intelligence has emerged as a promising tool and strategy for enhancing student well-being and safety, particularly in addressing the pervasive issue of cyberbullying. As educational institutions seek to leverage AI tools to create safer online environments and promote positive learning experiences, **it becomes imperative to evaluate the effectiveness, potential impact, and ethical implications of these technologies.** To this end, the evaluation schema delineates a structured framework comprising **three key components: capability, utility, and adoption.**

This division reflects the multi-layered nature of assessing AI technologies in ethically sensitive environments. **Each component serves a distinct yet interconnected indicator in evaluating the ethical deployment and practical efficacy of AI tools within the educational framework.**

By delineating these components, the evaluation schema provides a comprehensive framework for assessing the ethical use and practical impact of AI technologies in education. This structured approach enables educational stakeholders to systematically evaluate the technical capabilities, practical usefulness, and ethical considerations associated with AI deployment, thereby fostering informed decision-making and responsible integration of AI tools within educational ecosystems.

3.1. Capability

The capability component of the evaluation schema focuses on **assessing the technical effectiveness of AI systems deployed within educational environments.** This entails evaluating the AI's proficiency in detecting and effectively addressing instances of cyberbullying. Key factors under scrutiny include the accuracy, reliability, and overall performance of the AI algorithms in identifying and mitigating cyberbullying incidents. **This evaluation ensures that AI systems possess the requisite technical capabilities to function effectively** as tools for promoting online safety and well-being among students.

Furthermore, **the capability component examines the compatibility of AI systems with existing educational infrastructure and platforms.** This involves assessing whether AI technologies can seamlessly integrate with and complement the school's digital ecosystem. Compatibility considerations encompass aspects such as data interoperability, system scalability, and ease of implementation. **By evaluating the technical compatibility of AI systems, educational institutions can ensure smooth integration and optimal utilization** of these technologies within their existing frameworks.

3.2. Utility

The utility component of the evaluation schema **delves into the practical usefulness of AI applications in educational contexts.** This assessment focuses on **evaluating the tangible benefits and outcomes derived from the deployment of AI tools,** particularly in the realm of cyberbullying prevention and student well-being. Central to this evaluation is the examination of how effectively AI applications contribute to reducing instances of cyberbullying and fostering a safer online environment for students.



In addition to cyberbullying prevention, **the utility component also scrutinizes the broader educational impacts facilitated by AI technologies.** This includes assessing their efficacy in delivering personalized learning experiences, enhancing student engagement, and improving academic performance. By evaluating the utility of AI applications beyond cyberbullying prevention, educational stakeholders can gain insights into the holistic benefits and potential drawbacks of integrating AI into educational practices.

3.3. Adoption

The adoption component of the evaluation schema focuses on **assessing the extent to which AI technologies are embraced and integrated into educational practices.** This evaluation examines the level of acceptance and utilization of AI tools by educators, students, and other stakeholders within the school community. **It seeks to determine the degree of adoption of AI technologies and the factors influencing their uptake within educational settings.**

Moreover, **the adoption component measures also the alignment with EU values, ethical standards, and regulatory frameworks governing data privacy and protection.** This involves assessing whether AI applications adhere to principles of fairness, transparency, and accountability in their implementation and use. By evaluating the adoption of AI technologies within the context of ethical and regulatory considerations, educational institutions can ensure responsible and sustainable integration of these technologies into their practices.

4 Evaluation activities

THEORETICAL DIMENSION			
Data Privacy Compliance Evaluation Criteria			
Activity	Description	Expected Output	KPIs
Assessment of the minimum requirements in term of Data Privacy	This activity involves conducting a comprehensive assessment of the AI-based educational tool or practice to ensure compliance with minimum data privacy requirements. It includes evaluating data collection, storage, processing practices, and security measures implemented to protect student data.	A comprehensive report outlining the current state of data privacy within the AI tool or practice, highlighting areas of compliance and potential gaps or risks.	• Percentage of data privacy requirements met by the AI tool or practice in accordance with the higher standards of protection. • Number of identified data privacy risks or vulnerabilities. • Implementation status of recommended measures to address data privacy gaps.

Ethics Compliance Evaluation Criteria			
Activity	Description	Expected output	KPIs
Mapping and understanding of the potential elements of ethical risk	This activity involves mapping and analyzing potential ethical risks associated with the use of AI in educational settings. It includes identifying potential biases, discrimination, lack of transparency, or other ethical concerns inherent in the AI-based educational tool or practice.	A comprehensive ethical risk assessment report highlighting potential ethical dilemmas, biases, and risks associated with the AI tool or practice, along with recommendations for mitigation strategies.	• Number of identified ethical risks or concerns. • Severity rating of identified ethical risks. • Implementation status of recommended mitigation strategies.
EU Recommendation Evaluation Criteria			
Activity	Description	Expected output	KPIs
Benchmark between the implemented solutions and the last EU policy recommendations	This activity involves comparing the features and functionalities of the implemented AI solutions in educational settings with the latest policy recommendations provided by the European Union (EU). It includes assessing alignment with EU regulations, guidelines, and ethical principles.	A comparative analysis report highlighting the degree of alignment between the implemented AI solutions and the latest EU policy recommendations, identifying areas of compliance and potential areas for improvement.	• Percentage of alignment with EU policy recommendations. • Identification of areas where the implemented solutions deviate from EU recommendations. • Implementation status of recommendations to

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Project No: 2023-1-ES01-KA220-SCH-000160237

			enhance alignment with EU policies.
--	--	--	-------------------------------------

IMPACT DIMENSION			
Capability Evaluation Criteria			
Activity	Description	Expected Output	KPIs
Technical Evaluation of AI Systems	Conduct a comprehensive technical assessment of AI systems deployed in educational environments. This includes evaluating the algorithms, accuracy rates, and performance metrics of AI tools in detecting and addressing cyberbullying incidents.	Observation report outlining the technical capabilities and limitations of AI systems, including recommendations for improvement.	• Accuracy rate of AI algorithms in detecting cyberbullying incidents. • Performance metrics such as false positive and false negative rates. • Compatibility with existing educational infrastructure and platforms.

Utility Evaluation Criteria			
Activity	Description	Expected output	KPIs
Impact Evaluation of AI Applications	Assess the practical usefulness and impact of AI applications in educational contexts. This involves evaluating their effectiveness in promoting student well-being, reducing instances of cyberbullying, and enhancing educational outcomes.	Evaluation report highlighting the tangible benefits and outcomes derived from the deployment of AI tools, along with recommendations for optimizing their utility.	• Reduction in instances of cyberbullying following the implementation of AI applications. • Improvement in student well-being and safety. • Enhanced academic performance and engagement.
Adoption Evaluation Criteria			
Activity	Description	Expected output	KPIs
Stakeholder Perception Survey	Administer surveys to educators, students, and other stakeholders to assess their perceptions and acceptance of AI technologies in educational settings. This includes gauging attitudes towards AI adoption, identifying barriers to implementation, and soliciting feedback on usability.	Survey analysis report summarizing stakeholder perceptions and recommendations for promoting the adoption of AI technologies	• Level of acceptance and willingness to adopt AI technologies among educators and students. • Number of identified key barriers to adoption and implementation.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Project No: 2023-1-ES01-KA220-SCH-000160237

			• Feedback on usability and user experience of AI applications.
--	--	--	---

5 Conclusion and disclaimer

In conclusion, the Evaluation Schema for AI in Education, presented within this document, stands as a foundational framework pivotal for assessing the ethical, privacy, and compliance facets of AI-based educational tools and practices. **Beside its relevance, this document has some limits that have to be taken in consideration by the reader.**

While the Evaluation Schema for AI in Education endeavors to provide a comprehensive framework, it is essential to acknowledge the inherent complexity and diversity of the educational landscape. Recognizing that no single schema can fully encapsulate every nuance, institutions are encouraged to adopt a flexible approach. They should tailor and integrate the suggested evaluation activities to align with their unique contexts, institutional cultures, and the specific needs of their diverse target groups. By doing so, institutions can ensure that the evaluation process remains relevant, meaningful, and impactful within their particular educational environments.

Ethical considerations lie at the heart of any evaluation schema, given the inherently qualitative nature of ethical dilemmas. In the realm of AI in education, ethical scrutiny demands meticulous examination, taking into account potential biases, cultural nuances, and the evolving sensitivities of society. It necessitates a thorough exploration of the ethical implications inherent in the design, implementation, and utilization of AI technologies within educational settings. By engaging in thoughtful and reflective ethical inquiry, institutions can navigate complex ethical terrain, uphold principles of fairness and equity, and promote ethical decision-making practices that prioritize the well-being and rights of all stakeholders involved.

The impact evaluation outlined in this schema is designed to guide institutions towards cost-effective implementations, mindful of the resource constraints commonly faced in educational settings. However, institutions with the capacity to allocate additional resources are presented with opportunities for innovation and experimentation. By investing in AI technologies and educational practices, these institutions can explore novel approaches, pilot innovative solutions, and pioneer new methodologies that have the potential to significantly enhance the educational landscape. Through strategic investment and prudent resource allocation, institutions can unlock new possibilities, drive positive change, and foster a culture of innovation within the realm of AI in education.

It is crucial to recognize that this document represents an initial step in the ongoing journey of the AI@Mediators project. As technology and regulatory landscapes continue to evolve rapidly, stakeholders are urged to stay informed and remain abreast of project developments. Moreover, stakeholders are encouraged to reference other more recent project outputs, which may offer nuanced insights, updated guidance, and refined methodologies. By staying engaged and responsive to emerging trends and developments, stakeholders can ensure that their approaches to AI in education remain relevant, effective, and ethically sound, ultimately contributing to the advancement of educational practices and the well-being of learners worldwide.

6 References

- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(1), 38. <https://link.springer.com/article/10.1186/s41239-023-00408-3>
- JISC (2022) AI in tertiary Education. A summary of the current state of play. Retrieved from <https://repository.jisc.ac.uk/8783/1/ai-in-tertiary-education-report-june-2022.pdf>
- Martínez-Comesaña, M., Rigueira-Díaz, X., Larrañaga-Janeiro, A., Martínez-Torres, J., Ocaranza-Prado, I., & Kreibel, D. (2023). Impacto de la inteligencia artificial en los métodos de evaluación en la educación primaria y secundaria: revisión sistemática de la literatura. *Revista de Psicodidáctica*, 28(2), 93-103. <https://doi.org/10.1016/j.psicod.2023.06.001>
- UNESCO (2023). ChatGPT and artificial intelligence in higher education. Retrieved from https://www.iesalc.unesco.org/wp-content/uploads/2023/04/ChatGPT-and-Artificial-Intelligence-in-higher-education-Quick-Start-guide_EN_FINAL.pdf
- UNESCO (2023). Guidance for generative AI in education and research. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Van Laar, E., Van Deursen, A. J., Van Dijk, J. A., & De Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in human behavior*, 72, 577-588. <https://doi.org/10.1016/j.chb.2017.03.010>
- Hinduja, S., & Patchin, J. W. (2015). Cyberbullying prevention and response: Expert perspectives. Routledge. https://www.researchgate.net/publication/345269061_Patchin_J_W_Hinduja_S_Cyberbullying_Prevention_and_Response_Expert_Perspectives_New_York_Routledge_2012_204_pages
- Smith, P. K., & Steffgen, G. (Eds.). (2013). Cyberbullying through the new media: Findings from an international network. Psychology Press. <https://psycnet.apa.org/record/2013-37310-000>
- Cénat, J. M., Hébert, M., Blais, M., Lavoie, F., Guerrier, M., & Derivois, D. (2014). Cyberbullying, psychological distress and self-esteem among youth in Quebec schools. *Journal of affective disorders*, 169, 7-9. https://www.researchgate.net/publication/264084590_Cyberbullying_psychological_distress_and_self-esteem_among_youth_in_Quebec_Schools
- Juvonen, J., & Gross, E. F. (Eds.). (2008). Extending the school grounds?—Bullying experiences in cyberspace. *Journal of School Health*, 78(9), 496-505. https://www.researchgate.net/publication/23252828_Extending_the_School_Grounds-Bullying_Experiences_in_Cyberspace
- Patchin, J. W., & Hinduja, S. (2011). Cyberbullying prevention and response: Expert perspectives. Routledge. <https://www.routledge.com/Cyberbullying-Prevention-and-Response-Expert-Perspectives/Patchin-Hinduja/p/book/9780415892377>
- Belsey, B. (2004). Cyberbullying: An emerging threat to the “always on” generation. Cyberbullying Conference, 2-3. <https://billbelsey.com/?p=1827>
- Spears, B., Slee, P., Owens, L., & Johnson, B. (2009). Behind the scenes and screens: Insights into the human dimension of covert and cyberbullying. *Social Inclusion*, 35-47. https://www.researchgate.net/publication/232459474_Behind_the_Scenes_and_Screens_In_sights_into_the_Human_Dimension_of_Covert_and_Cyberbullying