



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 Policies and protocols for the use of AI in secondary educational and activity training field	3
1.1. What do we mean by reliable AI?	4
1.1.1. Harm Prevention Principle	4
1.1.2. Well-being principle	4
1.1.3. Respect for human autonomy	5
1.1.4. Non-discrimination	5
1.1.5. Transparency	6
1.1.6. Proactive responsibility	6
1.1.7. Justice	6
1.1.8. Explainability	7
1.1.9. Active participation	7
1.1.10. Privacy	8
1.1.11. Care for the environment	8
2 Bibliographical references	9
3 Annex	10

1 Policies and protocols for the use of AI in secondary educational and activity training field

Artificial Intelligence (AI from now on) has been the subject of debate in recent years due to the great impact it is having and will have on all areas of our lives. This guide focuses on a specific audience and context: teenagers and the educational context.

These elements are important for vulnerable minors who are exposed to the risk of AI tools usage. Nevertheless, at the same time they can benefit enormously from it (experimental examples can be found in Akour et al., 2023; Vivar & García-Peñalvo, 2023; Zhang & Chen, 2022). Some of the advantages that AI offers to young people are universal access, intelligent tutoring systems, virtual facilitators, intelligent content, collaboration between teachers and AI, content analysis, tutoring support outside the classroom, automation of administrative tasks, etc. Changes in education, because of its central role in a society, must be carefully observed and evaluated when they are introduced, as these changes can modify the behaviours and thinking of the population.

That is why this guide aims to provide teachers, pupils, educators and the general public with information and tools on the following topics:

- **What we mean by trustworthy AI:** A programme using AI should not only be useful to its users but also align with the values of Democracy and Human Rights.
- **What we should expect from the use of AI in education:** The use of AI can greatly improve efficiency in education, but to do so we must take steps to minimise risks and maximise its benefit.
- **How should AI be developed to be ethical and responsible?** A compilation of best practices for developing AI considering factors such as privacy, data security, potential biases, etc.
- **How AI should be implemented when used in education?** Factors to be considered when explaining its uses, possible risks, consent, the child's role, etc.

This guide will provide an explanation for the main problems associated with AI. Children are not only a high-risk factor because of their social status, but because of the crucial stage in their biological, psychological and social development as human beings.

The development of this guideline is based on a review of the current state of the art in European guidelines and protocols and peer-reviewed scientific work.

Still, the world of AI is maturing and its constant change makes us renew and evaluate every new piece of knowledge we acquire.

1.1. What do we mean by reliable AI?

The development of an AI, its purposes and how they are carried out, are contextualised by the system of interactions between individuals and the rules they agree to live by. In addition, we know that AI programmes can modify our behavior, and in the field of education, they can improve not only the way we teach, but also the way we co-exist in school. Therefore, all AI programmes must meet a series of requirements:

1.1.1. Harm Prevention Principle

The development and use of Artificial intelligence must not cause harm (psychological, physical, social, economic, etc.) to students, parents, teachers or any other social agent (European Commission et al., 2020). That is why the possible risks of its use must be eliminated or minimized. This principle is fundamental to protect adolescents from possible cases of cyberbullying, among others.

Different international organisations (UNESCO, 2021) include "Respect for Human Dignity" as a fundamental right. This intrinsic value available to minors, and in general to any individual, must not be repressed or endangered in the use and development of Artificial Intelligence applications (European Commission, 2019).

1.1.2. Well-being principle

The application of an AI tool should not only entail the absence of harm, but the improvement of the human condition, be it by developing the physical, mental, social and/or emotional aspect (European Commission et al., 2020). The consequences also have an impact on collective well-being, with improved levels of productivity (Miao et al., 2021), which in the academic context would translate into more effective forms of teaching, such as greater literacy (Vivar & García-Peñalvo, 2023).

The improvement of social relations is also considered, with special attention being paid to the fact that AI tools should enhance physical interaction and not only (Miao et al., 2021). There are elements that have gained prominence in recent years that can pose a danger to education such as plagiarism or cheating for exams and assignments (OECD, 2023). AI systems should be integrated into education to expand the individual's potentials and skills, facilitating more mechanical tasks and fostering cognitive skills.

There may be contradictory elements of inseparable benefit and harm at the same time. The principle of proportionality, which assesses the costs and benefits to humanity of these AI systems, is recognized. This assessment justifies their use if the following criteria are met (UNESCO 2021):

1. **Preponderance of social benefit:** The decision must generate a positive balance for society, where the benefits significantly outweigh the costs. This implies considering the well-being of all stakeholders, both in the short and long term.
2. **Respect for fundamental principles:** The decision must not violate any fundamental ethical or legal principles, such as justice, equality, freedom or human dignity. It is crucial to act with integrity and responsibility at all times.

Funded by the European Union. The opinions and points of view expressed in this website commit only the author(s) and not necessarily those of the European Union or of the Spanish Service for the Internationalisation of Education (SEPIE). Neither the European Union or the SEPIE National Agency can be considered responsible for them.

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

3. **Appropriateness and scientific rigor:** The method used to make the decision must be appropriate to the specific context and based on sound science. This implies a thorough analysis of available information, consultation with experts and the application of reliable methodologies.

1.1.3. Respect for human autonomy

The ability of individuals to act autonomously over their own actions, being able to make decisions without being conditioned by external factors and being responsible for them (European Commission, 2022). To this end, the use of AI must be based on an awareness of what uses can be made of the AI application and an elimination of possible biases or manipulations in its results that could condition the individual's will for the interests of third parties.

Respect for individual freedom and human autonomy on the part of AI systems must involve mitigating illegitimate coercion, risks to mental autonomy, manipulation and any threats that limit control over their lives. In any case, they should complement and enhance their cognitive, cultural and social skills as individuals.

Within the educational context, consideration should be given to which elements can be substituted or useful as support for teachers, and which elements should remain purely human (European Commission et al., 2020; OECD-Education International, 2023).

1.1.4. Non-discrimination

Another fundamental principle is the moral obligation to ensure that AI systems do not discriminate against these children based on (among others) race, nationality, gender, sexual orientation, religion or political beliefs (Slimi & Villarejo-Carballido, 2023; UNESCO, 2019). These variables can be collected as data for demographic studies, but never as a reason for exclusion in decisions that may affect the life of the individual (e.g. educational quality of the child).

AI systems are likely to have biases in their decisions that arise from human error, either because the human data entered is biased and/or because the algorithms applied and developed by humans are biased (Vivar & García-Peñalvo, 2023). The reasons for discrimination and its consequent marginalisation pose a serious mental and social health problem for both the excluded adolescent and the marginalised social group.

In their design, AI systems should aim to be accessible to the widest possible population, considering differences in language, age or possible disabilities (European Commission, 2022; Miao et al., 2021; UNESCO, 2021). This increased user access leads to more active participation in possible changes within and outside the AI system.

Extending to this principle, benefits and costs of AI use, including the creation of its infrastructure (technical equipment and connectivity) should be distributed fairly and equitably (OECD-Education International, 2023).

Funded by the European Union. The opinions and points of view expressed in this website commit only the author(s) and not necessarily those of the European Union or of the Spanish Service for the Internationalisation of Education (SEPIE). Neither the European Union or the SEPIE National Agency can be considered responsible for them.

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

1.1.5. Transparency

The most cited principle in all literature based on the use of AI and ethical principles (Jobin et al., 2019) as greater transparency contributes to more peaceful, just, democratic and inclusive societies by enabling public scrutiny and reducing corruption and discrimination (UNESCO, 2021). It is based on the assumption that users should be informed of clear and understandable ways about the use of AI and all that it entails (Nguyen et al., 2022; OECD-Education International, 2023):

That an AI application is being used

1. **What data are being collected and for what purpose**, including obtaining informed consent.
2. **Why it gives those results or makes those decisions**, while maintaining a balance between giving as much information as possible and being as understandable as possible for interested users.
3. **What are the risks and limitations of such AI**, in order to be able to assess the pros and cons of its use and to be able to interpret the extent to which its results are effective.

Data and processes for both decision-making and data collection and labelling must be documented, traceable and auditable in order to identify potential errors and prevent future ones (European Commission et al., 2020; European Commission et al., 2022; Miao et al., 2021; UNESCO, 2019).

1.1.6. Proactive responsibility

Focused on the use of data in AI applications, this refers to the ethical commitment required of social actors involved in the development, distribution and implementation of these applications (Miao et al., 2022). These social actors must anticipate in reducing or eliminating any potential risks that these technologies may pose to adolescents (European Commission, 2022) as an AI application alone cannot understand the biases or power differentials it may produce (John-Mathews, 2022).

It is therefore not only a matter of complying with high standards, but also of adopting a proactive stance to ensure the future well-being of children and adolescents. For true proactive responsibility to occur, all developers of AI systems must have a legal entity that is accountable for possible negative effects, as software cannot be taught ethics (UNESCO, 2021; Vivar & García-Peñalvo, 2023), and its technical side must be auditable by the competent authorities.

It is necessary that the moral involvement of companies and institutions materialize in greater social wellbeing and do not become an obstacle to the development of a new society, use as a tool to circumvent legal barriers or for public image issues (*ethical washing*) (Henry & Oliver, 2021).

1.1.7. Justice

This is the moral obligation to ensure that the development and use of AI is fair, impartial and does not discriminate against any student based on gender, race, religion or political beliefs (European Commission, 2022). This principle is fundamental to ensure the responsible and ethical use of Artificial Intelligence programmes, as well as ensuring that all children, guardians and teachers have equal

Funded by the European Union. The opinions and points of view expressed in this website commit only the author(s) and not necessarily those of the European Union or of the Spanish Service for the Internationalisation of Education (SEPIE). Neither the European Union or the SEPIE National Agency can be considered responsible for them.

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

access and benefit equally (OECD- Education International, 2023). The principle of Justice includes compliance with laws, the rule of law and democratic principles (European Commission, 2019).

1.1.8. Explainability

This concept gives meaning to informed consent because if the user does not know exactly what they are accepting, their rights are violated. It is therefore not a matter of giving all the information in technical and incomprehensible language, but of conveying the key points regarding its use or the collection of data, results and possible risks of using the application (European Commission et al., 2020). This is particularly important for minors, as they may not have the vocabulary and comprehension skills of an adult.

Explainability enables legal guardians and students to make informed decisions about their use and to identify and mitigate potential risks on the part of the users (UNESCO, 2021). Promoting an understandable explanation of AI use will encourage trust in the programme and thus more efficient use (Farrow, 2023). In this sense, teachers using AI software require time, professional development and working conditions, without the need to become a data scientist, to improve the understanding and effectiveness of these resources (Chounta et al., 2021; OECD-Education International, 2023).

At the technical level, special attention should be paid to so-called "*black box*" algorithms, which refer to the difficulty or impossibility of explaining why a model has generated a decision. Such algorithms, or their weight in the educational context, should be minimized or eliminated (European Commission, 2019; Farrow, 2023).

1.1.9. Active participation

Users of AI applications should not be seen as mere passive agents but need to be involved (or represented) in the design, development and use to avoid problems with the new scenarios posed by AI systems (European Parliamentary Research Service, 2023; Slimi & Villarejo- Carballido, 2023). This includes not only their involvement in the design and use of these tools, but also their auditability, minimisation and reporting of negative side effects, trade-offs and compensations (Nguyen et al., 2022). Understanding how AI applications work for children and their caregivers not only facilitates previously discussed principles such as explainability, but also the ability to tailor them to their needs to improve the effectiveness of the educational environment (UNESCO, 2019).

Similarly, teachers should have access to and be trained in quality content on the use of AI in the classroom to improve their instructional methods (Chounta et al., 2021; OECD-Education International, 2023; Lamas & Arnab, 2021). The interests of minors, and by extension their legal guardians, should be realised in the purposes of these artificial intelligences, not only by interfering with their wishes, but with the active collaboration of minors or their representatives (Miao et al., 2021). This participation indirectly helps children and adults understand AI applications and develop their digital skills and critical and creative thinking (OECD-Education International, 2023).

1.1.10. Privacy

It is the protection of personal data of individuals, specifically children and adolescents, from unauthorized collection, use and processing. This is particularly important when balancing privacy and open access to the AI system infrastructure and its effectiveness for the intended purpose (Lanna, 2022; Miao et al., 2021). AI tools must not only justify that the data they collect is legitimate, and for specific purposes consented to by the child and legal guardian.

They must also reassure that they do not collect other data, in particular biometric data, for which the user has not given informed consent or protect data that was consented to against unauthorised access, misuse, modification, disclosure or destruction (European Commission, 2022; OECD-Education International, 2023). There is a growing danger from AI systems that individuals are constantly being watched, exposed and judged by past or irrelevant elements (Vincent-Lancrin & Van Der Vlies, 2020). Minimising and anonymising the personal data of minors reduces potential risks to be taken into account (European Commission, 2019). This principle includes the right of access, rectification and deletion of the user's personal data at the will of the minor and/or the legal guardian (Article 8 - Protection of personal data, 2007).

1.1.11. Care for the environment

The use of artificial intelligence involves considerable energy expenditure (Cho, 2023), so developers and distributors must be responsible for using AI in the most sustainable and environmentally friendly way, minimising its environmental impact as well as maximising its cost-effectiveness (European Commission, 2022). Furthermore, given the character of existential necessity for humanity, the prosperity of the environment and ecosystems must be considered as part of the duties we have as a society, giving importance to the life cycles of AI systems for sustainable development (UNESCO, 2021). The social impact that AI systems can have should change based on how we interact and behave. One of the directions of this impact should be to promote awareness of sustainability and ecological responsibility.

The implementation of Artificial Intelligence (AI) in education presents significant opportunities to improve teaching and combat cyberbullying, but it also poses significant ethical and social challenges, especially when it comes to protecting adolescents. To ensure reliable and ethical AI in education, it is essential that social actors (students, teachers, developers, legal guardians, etc.) work together to develop these systems and follow the principles mentioned in this guide. By adhering to these ethical values, the responsible use of AI in education can be promoted, maximising its benefits while protecting the welfare and rights of students.

It is crucial to underline the novelty of the Artificial Intelligence (AI) paradigm in the educational context, which calls for a cautious approach and constant updating. The field is constantly evolving, requiring continuous vigilance to not only identify emerging tools and opportunities, but also to understand the potential limitations and risks it may present for younger learners.

2 Bibliographical references

NOTE: The underlined bibliographical references are the most relevant guidelines on the ethical use of artificial intelligence.

Akour, I.A., Muhammad H., Alzoubi, H. M., Alshurideh, M. T. & Ahmad Y. (2023) Integrating Artificial Intelligence in Improving Educational System: The Mediating role of Smart Learning. 2023 *International Conference on Business Analytics for Technology and Security (ICBATS)*, Dubai, United Arab Emirates, pp. 1-5, doi: 10.1109/ICBATS57792.2023.10111284.

Article 8 - Protection of personal data (2007, 14th of December). European Union Agency For Fundamental Rights. <https://fra.europa.eu/en/eu-charter/article/8-protection-personal-data>

Cho, R. (2023, 9 June). *AI's growing carbon footprint*. State Of The Planet. <https://news.climate.columbia.edu/2023/06/09/ais-growing-carbon-footprint/>

Chounta, I., Bardone, E., Raudsep, A., & Pedaste, M. (2021). Exploring Teachers' Perceptions of Artificial Intelligence as a Tool to Support their Practice in Estonian K-12 Education. *International Journal Of Artificial Intelligence In Education*, 32(3), 725-755. <https://doi.org/10.1007/s40593-021-00243-5>

European Commission (2019). *Ethics guidelines for trustworthy AI*, Publications Office. <https://data.europa.eu/doi/10.2759/346720>

European Commission (2022). *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for Educators*. Publications Office of the European Union. <https://www.internetsegura.pt/sites/default/files/2022-11/ethical-guidelines-on-the-use-of-artificial-intelligence-nc0722649enn.pdf>

European Commission, Joint Research Centre, Vuorikari, R., Punie, Y., Cabrera, M. (2020). *Emerging technologies and the teaching profession: ethical and pedagogical considerations based on near-future scenarios*, Publications Office. <https://data.europa.eu/doi/10.2760/46933>

European Parliamentary Research Service (2023). *Artificial intelligence act*. European Parliament. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)

Farrow, R. (2023). The possibilities and limits of XAI in education: a socio-technical perspective. *Learning, Media And Technology (Print)*, 48(2), 266-279. <https://doi.org/10.1080/17439884.2023.2185630>

Henry, J. V., & Oliver, M. (2021). Who Will Watch the Watchmen? The Ethico-political Arrangements of Algorithmic Proctoring for Academic Integrity. *Postdigital Science And Education*, 4(2), 330-353. <https://link.springer.com/article/10.1007/s42438-021-00273-1>

Jobin, A., Ienca, M. & Vayena, E. (2019). *The global landscape of AI ethics guidelines*. *Nature Machine Intelligence*, 1(9), 389-399. <https://doi.org/10.1038/s42256-019-0088-2>

Funded by the European Union. The opinions and points of view expressed in this website commit only the author(s) and not necessarily those of the European Union or of the Spanish Service for the Internationalisation of Education (SEPIE). Neither the European Union or the SEPIE National Agency can be considered responsible for them.

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

- John-Mathews, J. (2022). Some critical and ethical perspectives on the empirical turn of AI interpretability. *Technological Forecasting And Social Change*, 174, 121209. <https://doi.org/10.1016/j.techfore.2021.121209>
- Lameras, P., & Arnab, S. (2021). Power to the Teachers: An Exploratory Review on Artificial Intelligence in Education. *Information (Basel)*, 13(1), 14. <https://doi.org/10.3390/info13010014>
- Lanna, L. C. (2022). Literature review of the reciprocal value of artificial and human intelligence in early childhood education. *Journal Of Research On Technology In Education*, 55(1), 21-33. <https://doi.org/10.1080/15391523.2022.2128480>
- Miao, F., Holmes, W., Ronghuai, H. & Hui, Z. (2021). *AI and education Guidance for policy-makers*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000376709>
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. T. (2022). Ethical principles for artificial intelligence in education. *Education And Information Technologies*, 28(4), 4221-4241. <https://doi.org/10.1007/s10639-022-11316-w>
- OECD-Education International (2023) *Opportunities, Guidelines and Guardrails on Effective and Equitable Use of AI in Education*, OECD Publishing, Paris.
- Slimi, Z., & Villarejo-Carballido, B. (2023). Navigating the Ethical Challenges of Artificial Intelligence in Higher Education: An Analysis of Seven Global AI Ethics Policies. *TEM Journal*, 590-602. <https://doi.org/10.18421/tem122-02>
- UNESCO (2019). *Beijing Consensus on Artificial Intelligence and Education* [Press release]. <https://unesdoc.unesco.org/ark:/48223/pf0000368303>
- UNESCO (2021). *Recommendation on the Ethics of Artificial Intelligence* [Press release]. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- Vincent-Lancrin, S., & Van Der Vlies, R. (2020). Trustworthy artificial intelligence (AI) in education. *OECD Education Working Papers*. <https://doi.org/10.1787/a6c90fa9-en>
- Vivar, J. M. F., & García-Peñalvo, F. J. (2023). Reflections on the ethics, potential, and challenges of artificial intelligence in the framework of quality education (SDG4). *Comunicar (Huelva)*, 31(74), 37-47. <https://doi.org/10.3916/c74-2023-03>
- Zhang, S. & Chen, X. (2022) Applying Artificial Intelligence into Early Childhood Math Education: Lesson Design and Course Effect. In *2022 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)* (pp. 635- 638). [doi: 10.1109/TALE54877.2022.00109](https://doi.org/10.1109/TALE54877.2022.00109)

3 Annex

Read the detailed guide of the policies and protocols that concern the ethical use of Artificial Intelligence in (secondary) education: <https://drive.google.com/file/d/1Sz-IQOqY0FkQI-PoJdftQiPsjJmyNpEc/view?usp=sharing>

Funded by the European Union. The opinions and points of view expressed in this website commit only the author(s) and not necessarily those of the European Union or of the Spanish Service for the Internationalisation of Education (SEPIE). Neither the European Union or the SEPIE National Agency can be considered responsible for them.

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