Funded by the European Union

ARTIFICIAL INTELLIGENCE AND EDUCATION

Ethical issues on the use of AI in Education

EU Policy Context

U2SID - University to society collaborations for inclusive digital transformation in the Western Balkans

Inclusive Digital Transformation in Higher Education - U2SID Workshop.

Ph.D. MARSELA ROBO _NASRI

The EU Policy Context in AI Education

- In **2018** the **Commission** adopted a **communication to promote the development of AI in Europe**, and in **2019** it published a **coordinated plan on AI** – endorsed by the Council of the European Union – to coordinate the EU Member States' national AI strategies.
- Building on this groundwork, in **April 2019** the **Commission published a set of non-binding Ethics guidelines for trustworthy AI**. Prepared by the Commission's High-Level Expert Group on AI, composed of 52 independent experts, this document **aims to offer guidance on how to foster and secure the development of ethical AI systems in the EU**.
- The new **Digital Education Action Plan (2021-2027)** outlines the European Commission's vision for high-quality, inclusive and accessible digital education in Europe to foster the development of a high-performing digital education ecosystem and enhance digital skills and competences needed for digital transformation.
- As part of its digital strategy, the European Commission has proposed a comprehensive legal framework for AI (AI Act) laying down mandatory requirements for "high-risk" AI systems in several areas, including education and vocational training. This will build on the EU regulatory and policy developments on AI and data, which include the **Ethics Guidelines for Trustworthy AI** presented in **2019** by the High-Level Expert Group on AI (AI HLEG), the **General Data Protection Regulation (GDPR)** and the proposal for a **Data Act.**
- As AI systems increasingly affect all aspects of human activity, it seems essential that educators understand how to use AI tools in an ethical way in their teaching and their students' learning across the EU.
- Under Priority 1: Fostering the development of a high-performing digital education ecosystem, the Digital Education Action Plan outlines a set of actions to foster the development of a highperforming digital education ecosystem. This includes a specific action to develop ethical guidelines on the use of AI and data in education and training to be shared with educators and school leaders.
- To help tackle these challenges, in **October 2022** the Commission published **Ethical guidelines on the use of AI and data in teaching and learning for educators**.

ARTIFICIAL INTELLIGENCE AND EDUCATION

In 2019, the Council of Europe's Committee of Ministers adopted a recommendation on digital citizenship education in which a key focus was the **application of artificial intelligence (AI) in educational contexts**: *"AI, like any other tool, offers many **opportunities** but also carries with it many **threats**, which make it necessary to take human rights principles into account in the early design of its application. **Educators must be aware of the strengths and weaknesses of AI in learning, so as to be empowered – not overpowered – by technology in their digital citizenship education practices**. AI, via machine learning and deep learning, **can enrich education** … By the same token, developments in the AI field **can deeply impact interactions between educators and learners** and among citizens at large, which **may undermine the very core of education, that is, the fostering of free will and independent and critical thinking via learning opportunities** … Although it seems premature to make wider use of AI in learning environments, **professionals in education and school staff** should be made aware of AI and the **ethical challenges** it poses in the context of schools*. (Council of Europe 2019)

Defining AI

- **Oxford English Dictionary, 2006**: The capacity of computers or other machines to exhibit or simulate intelligent behaviour. Stanford University, 2016: A branch of computer science that studies the properties of intelligence by synthesizing intelligence.
- **OECD, 2019**: An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.
- **Council of Europe, 2021**: A set of sciences, theories and techniques whose purpose is to reproduce by a machine the cognitive abilities of a human being. Current developments aim, for instance, to be able to entrust a machine with complex tasks previously delegated to a human.
- **European Union, 2021**: Software that is developed with one or more of the techniques and approaches listed in Annex I *[(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods. and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with*].
- Parliamentary Assembly of the Council of Europe, 2021: Computer-based systems that can perceive and derive data from their environment, and then use statistical algorithms to process that data in order to produce results intended to achieve pre-determined goals. The algorithms consist of rules that may be established by human input, or set by the computer itself, which "trains" the algorithm by analysing massive data sets and continues to refine the rules as new data is received.
- **UNICEF, 2021:** Machine-based systems that can, given a set of human-defined objectives, make predictions, recommendations, or decisions that influence real or virtual environments. AI systems interact with us and act on our environment, either directly or indirectly. Often, they appear to operate autonomously, and can adapt their behaviour by learning about the context.
- **UNESCO, 2021:** AI systems are information-processing technologies that integrate models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments. AI systems are designed to operate with varying degrees of autonomy by means of knowledge modelling and representation and by exploiting data and calculating correlations.

Ethical considerations for AI and data in education and training

The ethics of AI is not only about **how we use technology**, but it is also about **why and for what reason** we use AI and data to support teaching, learning and assessment**.**

- Four different but interdependent ethical considerations that need to be taken into account when using AI and data in education and training: **agency, social fairness, humanity and justified choice.**
- The developmental dimension of education is centrally concerned with the question of how a person becomes a competent actor in a social and increasingly technology-mediated world. This question is approached using the concept of **agency.**
- The distribution of rights, responsibilities, resources, and power among different social groups generates a social dimension in ethics. The ethical challenges related to this are captured in the concept of social **fairness.**
- Ethics is also about what it is to be a human. To be an ethical human requires that we respect others and try to understand the world through the eyes of others. This implies openness, humility, and willingness to listen and learn. It also implies that all humans have dignity and cannot be treated as a means to an end. This aspect of ethics is about **humanity**.
- Ethics is deeply linked with our ideas on what counts as knowledge, fact, and evidence, and how these can be used to justify arguments. This aspect of ethics is concerned with questions on an acceptable argument, the concept of rationality, and the processes that are used to negotiate different points of view and different value systems. This requires transparency and leads to participatory and collaborative models of decision-making. This is captured with the concept of **justified choice**.

AI and Data Use Examples in Education

- The use of AI systems in classrooms across Europe is growing and AI is being used in different ways to support teaching, learning and assessment practices.
- **Ethical guidelines on the use of AI and data in teaching and learning for educators** provides four use-cases which are categorised as:

✓ Student Teaching – Using AI to teach students (student-facing);
✓ Student Supporting - Using AI to support student learning (student-facing);
✓ Teacher Supporting – Using AI to support the teacher (teacher-facing);
✓ System Supporting – Using AI to support diagnostic or system-wide planning (system-facing)

STUDENT TEACHING
Using AI to teach students

Intelligent tutoring system	The learner follows a step-by-step sequence of tasks and gets individualised instruction or feedback without requiring intervention from the teacher.
Dialogue-based tutoring systems	The learner follows a step-by-step sequence of tasks through conversation in natural language. More advanced systems can automatically adapt to the level of engagement to keep the learner motivated and on task.
Language learning applications	AI-based learning apps are used in formal and non-formal education contexts. They support learning by providing access to language courses, dictionaries and provide real-time automated feedback on pronunciation, comprehension and fluency.

STUDENT SUPPORTING
Using AI to support student learning

Exploratory learning environments	Learners are offered multiple representations that help them identify their own routes to achieving the learning goals.
Formative writing assessment	Learners are provided with regular automatic feedback on their writing/assignments.
AI-supported collaborative learning	Data on each learner's work style and past performance is used to divide them into groups with the same ability levels or suitable mix of abilities and talents. AI systems provide inputs/suggestions on how a group is working together by monitoring the level of interaction between group members.

TEACHER SUPPORTING
Using AI to support the teacher

Summative writing assessment, essay scoring	AI is used to evaluate and grade learners' written work automatically. AI and machine learning techniques identify features such as word usage, grammar and sentence structure to grade and provide feedback.
Student forum monitoring	Key words in student forum posts trigger automatic feedback. Discussion analytics provide insights to students' forum activity and can highlight students who may need help or are not participating as expected.
AI teaching assistants	AI agents or chatbots provide answers to commonly asked questions by learners with simple instruction and directions. Over time, the AI system is able to broaden the range of answers and options provided.
Pedagogical resource recommendation	AI recommendation engines are used to recommend specific learning activities or resources based on each student's preferences, progress and needs.

SYSTEM SUPPORTING
AI to support diagnostic or system-wide planning

Educational data mining for resource allocation	Schools gather student data which are analysed and used to plan how available resources can be best allocated for tasks like creating class groupings, assigning teachers, timetabling, and highlighting students who may require additional learning support.
Diagnosing learning difficulties	Using learning analytics, cognitive skills such as vocabulary, listening, spatial reasoning, problem-solving, and memory are measured and used to diagnose learning difficulties, including underlying issues that are hard for a teacher to pick up but might be detected early using AI systems.
Guidance services	AI based guidance services provide ongoing prompts or choice to create pathways for future education. Users can form a competence profile including previous education and include their own interests. From this data, combined with up-to-date course catalogue or study opportunity information, relevant study recommendations can be created using natural language processing.

Key Requirements for Trustworthy AI - are recommendable for any AI system deployed and used in education. They address important concerns, such as **the risk of bias or error affecting educational outcomes**:

- **Human agency and oversight** including fundamental rights, children's rights, human agency, and human oversight.
- **Questions for Educators:**

1 Human Agency and Oversight

- Is the teacher role clearly defined so as to ensure that there is a teacher in the loop while the AI system is being used? How does the AI system affect the didactical role of the teacher?
- Are the decisions that impact students conducted with teacher agency and is the teacher able to notice anomalies or possible discrimination?
- Are procedures in place for teachers to monitor and intervene, for example in situations where empathy is required when dealing with learners or parents?
- Is there a mechanism for learners to opt-out if concerns have not been adequately addressed?
- Are there monitoring systems in place to prevent overconfidence in or overreliance on the AI system?
- Do teachers and school leaders have all the training and information needed to effectively use the system and ensure it is safe and does not cause harms or violate rights of students?

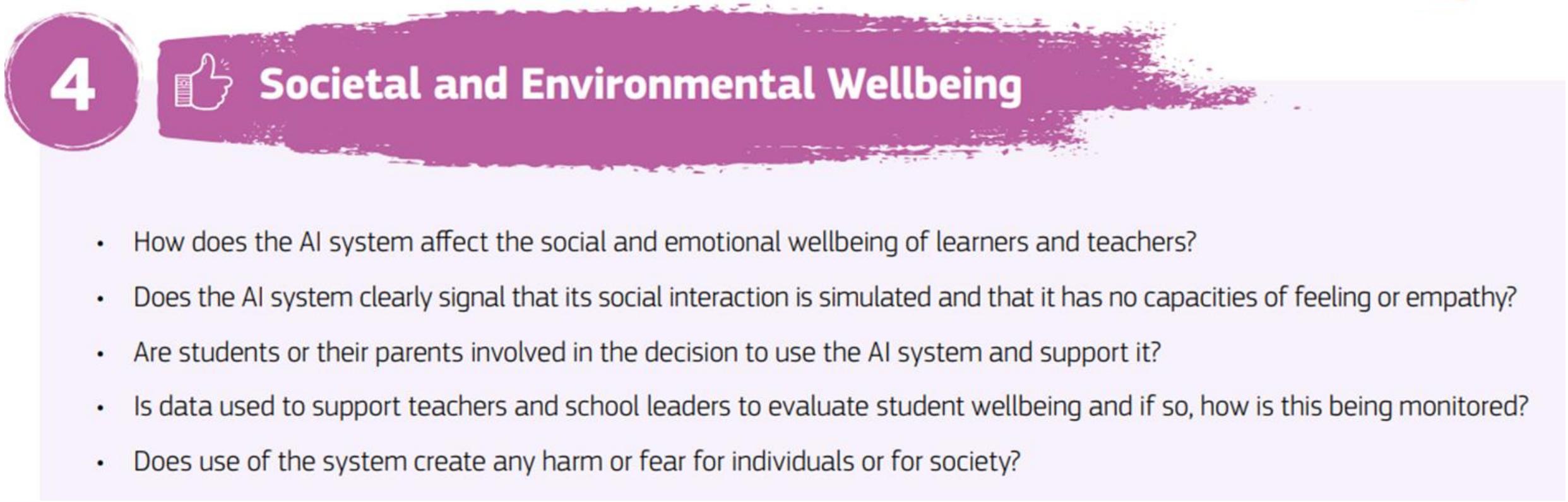
- **Transparency** including traceability, explainability and communication.
- **Questions for Educators:**

- Are teachers and school leaders aware of the AI methods and features being utilised by the system?
- Is it clear what aspects AI can take over and what not within the system?
- Do teachers and school leaders understand how specific assessment or personalisation algorithms work within the AI system?
- Are the system processes and outcomes focussed on the expected learning outcomes for the learners? How reliable are the predictions, assessments and classifications of the AI system in explaining and evaluating the relevance of its use?
- Are the instructions and information accessible and presented in a way that is clear both for teachers and learners?

- **Diversity, non-discrimination and fairness** including accessibility, universal design, the avoidance of unfair bias, and stakeholder participation, which allows use regardless of age, gender, abilities, or characteristics - with a particular focus for students with special needs.
- **Questions for Educators:**

- Is the system accessible by everyone in the same way without any barriers?
- Does the system provide appropriate interaction modes for learners with disabilities or special education needs? Is the AI system designed to treat learners respectfully adapting to their individual needs?
- Is the user interface appropriate and accessible for the age level of the learners? Has the usability and user-experience been tested for the target age group?
- Are there procedures in place to ensure that AI use will not lead to discrimination or unfair behaviour for all users?
- Does the AI system documentation or its training process provide insight into potential bias in the data?
- Are procedures in place to detect and deal with bias or perceived inequalities that may arise?

- **Societal and environmental wellbeing** including sustainability and environmental friendliness, social impact, society, and democracy.
- **Questions for Educators:**



- **Privacy and data governance** including respect for privacy, quality and integrity of data, and access to data.
- **Questions for Educators:**

- **Technical robustness and safety** including resilience to attack, security and general safety, accuracy, reliability, and reproducibility.
- **Questions for Educators:**

- Is there sufficient security in place to protect against data breaches?
- Is there a strategy to monitor and test if the AI system is meeting the goals, purposes and intended applications?
- Are the appropriate oversight mechanisms in place for data collection, storage, processing, minimisation and use?
- Is information available to assure learners and parents of the system's technical robustness and safety?

- **Accountability** including auditability, minimisation and reporting of negative impact, trade-offs, and redress. The considerations and requirements can help educators, school leaders and technology providers to adequately assess the impact, address the potential risks, and realise the benefits of an AI system deployed and used in education. As such they guide the development, deployment and use of trustworthy AI systems.
- **Questions for Educators:**

- Who is responsible for the ongoing monitoring of results produced by the AI system and how the results are being used to enhance teaching, learning and assessment?
- How is the effectiveness and impact of the AI system being evaluated and how does this evaluation consider key values of education?
- Who is responsible and accountable for final decisions made regarding the procurement and implementation of the AI system?
- Is there a Service Level Agreement in place, clearly outlining the support and maintenance services and steps to be taken to address reported problems?

Emerging Competences for Ethical use of AI and data

- Educators and school leaders play a central role in the successful adoption of AI systems and in realising the potential benefits of digital data in education.
- This will lead to the development of new digital competences to be considered in the context of the European Framework for the Digital Competence of Educators (DigCompEdu) which provides a general reference framework to support the development of educator-specific digital competences in Europe.
- Some potential indicators of the emerging educator and school leader competences for ethical use of AI and data in teaching and learning:

- Area 1: Professional Engagement Using digital technologies for communication, collaboration, and professional development
- Area 2: Digital resources Sourcing, creating, and sharing digital resources
- Area 3: Teaching and Learning Managing and orchestrating the use of digital technologies in teaching and learning.
- Area 4: Assessment Using digital technologies and strategies to enhance assessment.
- Area 5: Empowering Learners Using digital technologies to enhance inclusion, personalisation, and learners' active engagement.
- Area 6: Facilitating learners' digital competence Enabling learners to creatively and responsibly use digital technologies for information, communication, content creation, wellbeing and problem-solving

Area 1: Professional Engagement

Using digital technologies for communication, collaboration, and professional development

Competence element	Potential Indicators
Is able to critically describe positive and negative impacts of AI and data use in education	• Takes an active part in continuous professional learning on AI and learning analytics and their ethical use. • Able to give examples of AI systems and describe their relevance. • Knows how the ethical impact of AI systems is assessed in the school. • Knows how to initiate and promote strategies across the school and its wider community that promote ethical and responsible use of AI and data.
Understand the basics of AI and learning analytics	• Aware that AI algorithms work in ways that are usually not visible or easily understood by users. • Able to interact and give feedback to the AI system to influence what it recommends next. • Aware that sensors used in many digital technologies and applications generate large amounts of data, including personal data, that can be used to train an AI system. • Aware of EU AI ethics guidelines and self-assessment instruments.

Area 2: Digital resources

Sourcing, creating, and sharing digital resources

Competence element	Potential Indicators
Data governance	• Aware of the various forms of personal data used in education and training. • Aware of responsibilities in maintaining data security and privacy. • Knows that the processing of personal data is subject to national and EU regulation including GDPR. • Knows that processing of personal data usually cannot be based on user consent in compulsory education. • Knows who has access to student data, how access is monitored, and how long data are retained. • Knows that all EU citizens have the right to not be subject to fully automated decision making. • Able to give examples of sensitive data, including biometric data. • Able to weigh the benefits and risks before allowing third parties to process personal data especially when using AI systems.
AI governance	• Knows that AI systems are subject to national and EU regulation (notably AI Act to be adopted). • Able to explain the risk-based approach of the AI Act (to be adopted). • Knows the high-risk AI use cases in education and the associated requirements under the AI Act (when adopted). • Knows how to incorporate AI edited/manipulated digital content in one's own work and how that work should be credited. • Able to explain key principles of data quality in AI systems.

Area 3: Teaching and Learning

Managing and orchestrating the use of digital technologies in teaching and learning

Competence element	Potential Indicators
Models of learning	• Knows that AI systems implement designer's understanding of what learning is and how learning can be measured; can explain key pedagogic assumptions that underpin a given digital learning system.
Objectives of education	• Knows how a given digital system addresses the different social objectives of education (qualification, socialisation, subjectification).
Human agency	• Able to consider the AI system impact on teacher autonomy, professional development, and educational innovation. • Considers the sources of unacceptable bias in data-driven AI.
Fairness	• Considers risks related to emotional dependency and student self-image when using interactive AI systems and learning analytics.
Humanity	• Able to consider the impact of AI and data use on the student community. • Confident in discussing the ethical aspects of AI, and how they influence the way technology is used.
Participates in the development of learning practices that use AI and data	• Can explain how ethical principles and values are considered and negotiated in co-design and co-creation of learning practices that use AI and data (linked to learning design).

Area 4: Assessment

Using digital technologies and strategies to enhance assessment

Competence element	Potential Indicators
Personal differences	• Aware that students react in different ways to automated feedback.
Algorithmic bias	• Considers the sources of unacceptable bias in AI systems and how it can be mitigated.
Cognitive focus	• Aware that AI systems assess student progress based on pre-defined domain-specific models of knowledge. • Aware that most AI systems do not assess collaboration, social competences, or creativity.
New ways to misuse technology	• Aware of common ways to manipulate AI-based assessment.

Area 5: Empowering Learners

Using digital technologies to enhance inclusion, personalisation, and learners' active engagement

Competence element	Potential Indicators
AI addressing learners' diverse learning needs	• Knows the different ways personalised learning systems can adapt their behaviour (content, learning path, pedagogical approach). • Able to explain how a given system can benefit all students, independent of their cognitive, cultural, economic, or physical differences. • Aware that digital learning systems treat different student groups differently. • Able to consider impact on the development of student self-efficiency, self-image, mindset, and cognitive and affective self-regulation skills.
Justified choice	• Knows that AI and data use may benefit some learners more than others. • Able to explain what evidence has been used to justify the deployment of a given AI system in the classroom. • Recognises the need for constant monitoring of the outcomes of AI use and to learn from unexpected outcomes.

Area 6: Facilitating learners' digital competence

Enabling learners to creatively and responsibly use digital technologies for information, communication, content creation, wellbeing and problem-solving

Competence element	Potential Indicators
AI and Learning Analytics ethics	• Able to use AI projects and deployments to help students learn about ethics of AI and data use in education and training.