AI Ethical Frameworks

Lidija Kralj, Educational Analyst and Advisor
lidija.kralj@ucitelji.hr

EduCon

Digital Services Act

Article 28 Online protection of minors
1. Providers of online platforms accessible to minors shall put in place appropriate and proportionate measures to ensure a high level of privacy, safety, and security of minors, on their service.

Recital 70 Consequently, online platforms should consistently ensure that recipients of their service are appropriately informed about how recommender systems impact the way information is displayed, and can influence how information is presented to them.

EduCon

Digital Services Act

Algorithms for adapting task complexity or recommending learning content

Capture attention and keep users engaged within the platform

Reward systems

It would be a huge benefit for all education stakeholders if providers became more transparent and describe such systems and algorithms. Knowing more about how educational platforms work could help students gain a better understanding of their learning processes and at the same time give them the necessary agency.

Cyber Security Act

- The major aims of the Cyber Security Act (CSA) are achieving a high level of cybersecurity, cyber resilience and trust within the Union and increase citizens', organisations' and businesses' awareness of cybersecurity issues by organising the European Union Agency for Cybersecurity (ENISA) and offering information in a transparent manner on the level of security of ICT products, ICT services and ICT processes through creating a framework for the establishment of European cybersecurity certification schemes.
- CSA explicitly mention education, even dedicate a separate article to Awareness-raising and education. Article 10 ENISA shall: (a) raise public awareness of cybersecurity risks, and provide guidance on good practices for individual users aimed at citizens, organisations and businesses, including cyber-hygiene and cyber-literacy; …

EduCon

Data Act & Data Governance Act

Possible impacts on education:

- **Enhanced Data Access** because educational institutions often require access to diverse datasets for research and development, which can lead to improved access to educational resources and research data.
- **Encourages data sharing across** sectors, which can foster collaboration between educational institutions and other entities. **Data altruism**
- Requires **clear and transparent communication** about data practices. By sharing best practices and case studies, institutions can learn from each other's experiences and avoid common pitfalls. This can include successful strategies for data protection, innovative uses of data in education, and effective change management approaches.
- Promotion of **interoperability and standardization** of data formats as well as data altruism.

EduCon

Artificial Intelligence Act (AI Act)

- Emphasises the significance of utilising AI systems to update education systems, improve educational standards in both offline and online settings, and expand access to digital education for a broader range of people.
- Nevertheless, the utilisation of AI systems in the field of education, namely for making judgements related to admissions, evaluations, and selecting suitable educational levels, raises ethical concerns.
- These AI systems should be categorised as high-risk because of their capacity to influence an individual's educational and professional path, so affecting their ability to secure their means of living.
- Inadequately designed and utilised AI systems can intrude upon privacy, infringing upon the right to education, perpetuating prejudice, and fortifying longstanding prejudices against particular groups, such as women, specific age cohorts, individuals with disabilities, or those of specific racial, ethnic, or sexual orientations.

AI Act - Article 5 Prohibited AI practices

- (a)... AI system that deploys **subliminal techniques** beyond a person's consciousness or purposefully manipulative or deceptive techniques,...
- (b) ... AI system that exploits any of the **vulnerabilities** **of a natural person** or a specific group of persons due to their age, disability or a specific social or economic situation...
- (c) the placing on the market, the putting into service or the use of AI systems for the **evaluation or classification** of natural persons or groups of persons over a certain period of time based on their social behaviour or known, inferred or predicted personal or personality characteristics, with the social score leading to either or both of the following:
    - (i) detrimental or unfavourable treatment of certain natural persons or groups of persons in social contexts that are unrelated to the contexts in which the data was originally generated or collected;
    - (ii) detrimental or unfavourable treatment of certain natural persons or groups of persons that is unjustified or disproportionate to their social behaviour or its gravity;
- (f) ... AI systems to **infer emotions** of a natural person in the areas of workplace and education institutions, ...
- (g)... of **biometric categorisation systems** that categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation...

ANNEX III High-risk AI systems referred to in Article 6(2)

3. Education and vocational training:

- *(a) AI systems intended to be used to **determine access or admission** or to assign natural persons to educational and vocational training institutions at all levels;*
- *(b) AI systems intended to be used to **evaluate learning outcomes**, including when those outcomes are used to steer the learning process of natural persons in educational and vocational training institutions at all levels;*
- *(c) AI systems intended to be used for the purpose of **assessing the appropriate level of education** that an individual will receive or will be able to access, in the context of or within educational and vocational training institutions at all levels;*
- *(d) AI systems intended to be used **for monitoring and detecting prohibited behaviour** of students during tests in the context of or within educational and vocational training institutions at all levels.*

Possible scenarios

- AI systems used to determine access or admission analyse applications and make decisions on student admissions
- AI systems that evaluate learning outcomes, such as automated grading tools.
- AI systems used to monitor and detect prohibited behaviour during exams or in classrooms.
- AI-driven personalised learning platforms that adapt educational content based on individual student performance and behaviour.
- AI systems that predict student performance and provide insights into potential academic outcomes.

EduCon

Maybe not high-risk AI systems (Article 6)

- various situations in education which do not pose a significant threat to the health, safety, or fundamental rights of individuals, and they do not substantially impact decision-making outcomes.
- narrow procedural task - sorting files, calculating average, identifying identical copies
- improve the result of a previously completed human activity - language or grammar improvement, document styling if the document is previously written by human
- detecting decision-making patterns - comparing results of human grading with some standard grading patterns, for example in national or maturity exams
- preparatory task for an assessment - translation, linking or referencing to other data sources

Assess all AI systems intended to be used in education (criteria in Article 7)

(a) the intended purpose of the AI system; Is it appropriate for education
(c) the nature and amount of the data processed and used by the AI system, in particular whether special categories of personal data are processed; What kind of data is collected from minors, and what are those data used for?
(d) the extent to which the AI system acts autonomously and the possibility for a human to override a decision or recommendations that may lead to potential harm; Does a teacher have some option to manually adjust personalisation algorithm?
(g) the extent to which persons who are potentially harmed or suffer an adverse impact are dependent on the outcome produced with an AI system, in particular because for practical or legal reasons it is not reasonably possible to opt-out from that outcome; Students usually doesn't have opt-out option for any activity in education, and they very much

Assess all AI systems intended to be used in education (criteria in Article 7)

(h) the extent to which there is an imbalance of power, or the persons who are potentially harmed or suffer an adverse impact are in a vulnerable position in relation to the deployer of an AI system, in particular due to status, authority, knowledge, economic or social circumstances, or age; Students might be in vulnerable position in education

(i) the extent to which the outcome produced involving an AI system is easily corrigible or reversible, taking into account the technical solutions available to correct or reverse it, whereby outcomes having an adverse impact on health, safety or fundamental rights, shall not be considered to be easily corrigible or reversible; How teachers may correct or change AI system outcome?

(j) the magnitude and likelihood of benefit of the deployment of the AI system for individuals, groups, or society at large, including possible improvements in product safety; Is it helpful for students? Is it better than some other pedagogical approach?

EduCon

Article 86 Right to explanation of individual decision-making

- *Any affected person subject to a decision …. shall have the **right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken**.*
- AI Act points out that high-risk AI systems shall be designed, that they can be effectively overseen by natural persons. **It is imperative to have a "human in the loop" in all applications of AI systems in education.**
- Citizens (students & teachers) need to know their rights

EduCon

I do know my rights.

Slido.com 2327561

slido

Please download and install the Slido app on all computers you use

I do know my rights

ⓘ Start presenting to display the poll results on this slide.

AI’s implications for education

- Using AI to tackle educational challenges - use of AI tools for human learning, for instance, to provide personalized learning experiences with intelligent tutoring systems and adaptive learning platforms.
- Educating about AI - teaching teachers about AI technologies, their uses, and their ethical implications
- Innovating education systems, involves rethinking how we teach and learn in light of AI’s growing influence.

Cukurova, M., Kralj, L., Hertz, B. & Saltidou, E. (2024). Professional Development for Teachers in the Age of AI. European Schoolnet. Brussels, Belgium., 2024

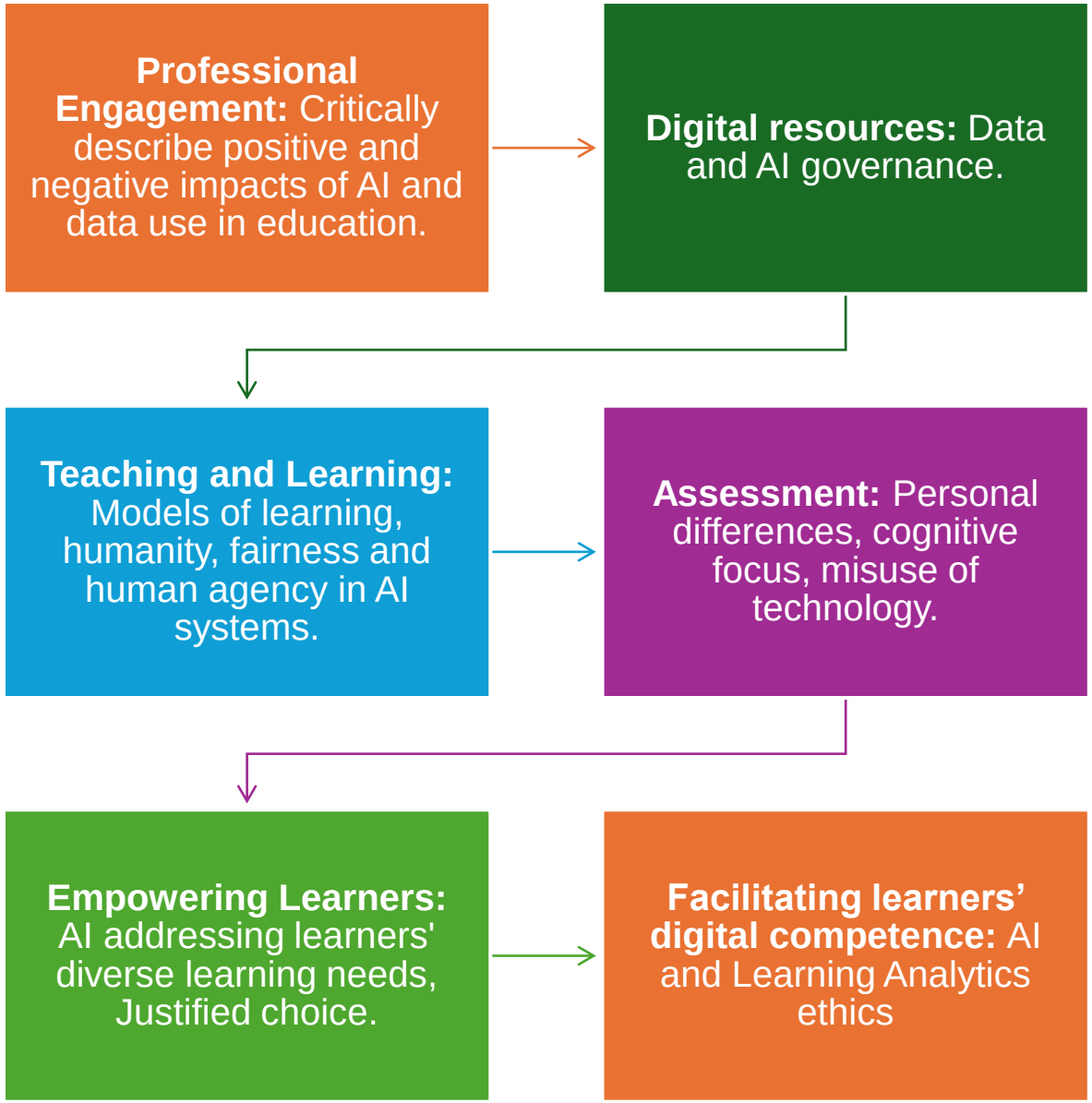
EduCon

What teacher training should focus on

- When comparing the requirements for teachers' competences in older frameworks that focus on ICT and digital technologies with newer frameworks that focus on the implementation of artificial intelligence technologies in education, it becomes evident that ethical considerations, human agency, and fairness are now given significant emphasis.
- In contrast, these aspects were previously assumed and not explicitly mentioned in older frameworks.
- One possible explanation for this phenomenon could be that the teaching profession places significant emphasis on ethics, human agency, and fairness, which were not jeopardised by information and communication technologies at the time but could be strongly negatively impacted by artificial intelligence today.

EduCon

European Commission's Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for Educators (2022)



EduCon

Ethics and humanity

Human agency encompasses an individual's capability to evolve into a proficient member of society, enabling them to make informed choices about their educational path and assume responsibility for their actions.

Fairness in education involves treating everyone fairly, ensuring equal access to opportunities through equity, inclusion, non-discrimination, and fair distribution of rights and responsibilities.

Humanity prioritizes people's identity, integrity, and dignity, focusing on well-being, safety, social cohesion, meaningful educational interactions, and respect.

Justified choice in educational settings involves using knowledge, facts, and data to justify collective decisions by stakeholders. This requires transparency in educational processes, participatory decision-making in policies, and the ability to provide explanations for educational choices and outcomes.

European Commission's Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for Educators (2022)

EduCon

Some examples of competences

1. Promoting and recommend-ing ethics as one of the core pillars when deploying AI systems in education in community.
2. Explaining how ethical principles and values are considered and negotiated in co-design and co-creation of learning practices that use AI and data
3. Actively contributing to the improvement of AI systems by reporting errors, risks, biases or misconcept-ions in data or outputs.
4. Comparing ethical principles built in AI tool with ethical principles teacher implement in teaching process.
5. Recognizing the high-risk AI use cases in education under the national or international regulations.
6. Advocating for the ethical use of AI tools, and actively participate in the development of a more ethically sound AI ecosystem in the field of education.

Please visit MENTI.COM 2101 5550

EduCon

Digital pedagogy

- When we look at the use of artificial intelligence in education it is necessary to first answer questions **why** we would like to use such technology; is the use of that AI technology the **best way** to support teaching and learning processes, **how** we will use it to support efficient achievement of learning outcomes; is it **appropriate** for our students and educational context; will it ensure **equitable use** for all students; do we have all **necessary consents** and licences to use it.

Some examples of competences

Discussing the best methods and criteria used for analysing and evaluating the AI tools, and their suitability for diverse users with peers.

Evaluating existing AI tools and resources for teaching and learning purposes.

Critically assessing AI's role in teaching and learning processes and support those findings with arguments and evidence.

Employing pedagogical methods in implementations of AI, ensuring a harmonious blend of human and AI supported teaching and learning.

Responsibly using AI tools and resources to enhance teaching effectiveness, efficiency, and differentiation.

Organizing collegial observations and debriefings of some AI supported educational activity to collect students' and colleagues' impressions and feedback.

Explaining key pedagogic assumptions that underpin a given AI learning system.

Please visit MENTI.COM 2101 5550

EduCon

How AI works

- Areas of **AI fundamentals** that teachers should be familiar with are probabilistic and statistical models which are base for more complex AI models, how automatization and decisions processes works and how AI systems use data.
- Learning analytics could offer valuable insights into teaching practices, enabling teachers to improve methods and strategies, identify struggling students, and provide necessary support. Data analytics and data visualisation are very important areas in the use of different AI tools, as teachers should be able to understand what AI tool is providing for them, usually as part of teacher's dashboards, what data is AI analysing, what conclusion could teachers make upon presented data visualisations.

EduCon

Some examples of competences

Demonstrating knowledge of good practices of protecting data and privacy when using AI tools.

Identifying and discussing the presence and impact of AI in educational context.

Recognizing various sources of bias in AI, from human inputs, data sets, or algorithms and understanding how automatic decision-making can be biased.

Explaining how a given system can benefit all students, independent of their cognitive, cultural, economic, or physical differences.

Researching about some of AI tools which are created for education, reading all documentations about it, contacting developers, and asking how their model is created and monitored.

Engaging in collaborative processes co-designing new products based on AI systems to support and enhance learning and teaching.

Being aware that AI is a rapidly changing area whose development and impact on education still remain unpredictable.

EduCon

Overview of some competency frameworks for teachers

- The European Framework for the Digital Competence of Educators (DigCompEdu), 2017
- The UNESCO ICT Competency Framework for Teachers (ICT-CFT) 2018
- Based on DigComEdu SELFIE FOR TEACHERS was developed as a self-reflection tool to support teachers in further developing their digital competence, 2021
- DigComp 2.2 (The Digital Competence Framework for Citizens) with new examples of knowledge, skills and attitudes, 2022
- Emerging Competences for Ethical use of AI and data from European Commission's Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for Educators (2022)
- Australian Education Ministry, Draft of National AI in Schools Framework 2023
- OECD Employment Outlook 2023 presents the different types of skills for the workforce that are becoming more prevalent because of AI.
- UNESCO (2023, 2024) draft Competency Frameworks for Teachers

EduCon

The European Framework for the Digital Competence of Educators (DigCompEdu, 2017)

- Professional Engagement area focuses on using digital technologies to enhance organizational communication with learners, parents, and third parties. It also includes professional collaboration, reflective practice, and digital continuous professional development.
- **Digital Resources** area involves identifying, assessing, and selecting digital resources for teaching and learning. It also includes modifying and building on existing openly-licensed resources and other resources where this is permitted, creating or co-creating new digital educational resources, and managing, protecting, and sharing digital resources.
- **Teaching and Learning** area covers planning for and implementing digital devices and resources in the teaching process, as to enhance the effectiveness of teaching interventions. It also includes using digital technologies to enhance the interaction with learners, individually and collectively, within and outside the learning session, and using digital technologies to foster and enhance learner collaboration.
- **Assessment** area involves using digital technologies to enhance formative and summative assessment, provide feedback, and track learner progress.
- **Empowering Learners** area focuses on using digital technologies to foster learner autonomy, scaffold mature use of digital technologies, differentiate teaching and learning, and promote digital inclusion.
- **Facilitating Learners' Digital Competence** area covers guiding learners in their development of digital competence, fostering responsible and safe use of digital technologies, and promoting learners' critical and creative thinking.

EduCon

DigComp 2.2 The Digital Competence Framework for Citizens 2022

- More than 30 examples of AI competences and an appendix describing 73 examples of citizens' competence when interacting with AI systems, like: What do AI systems do and what do they not do? How do AI systems work? The challenges and ethics of AI; Attitudes regarding human agency and control.
- Some of the AI examples from DigComp 2.2 applicable in educational context are:
- **Information and data literacy**: Weighs the benefits and disadvantages of using AI-driven search engines. Able to recognise that some AI algorithms may reinforce existing views in digital environments by creating “echo chambers” or “filter bubbles”.
- **Communication and collaboration:** Knows how to identify signs that indicate whether one is communicating with a human or an AI-based conversational agent. Recognises that AI that directly interacts with humans and takes decisions about their life can often be controversial.

EduCon

DigComp 2.2 The Digital Competence Framework for Citizens 2022

- **Digital content creation:** Knows that AI systems can be used to automatically create digital content using existing digital content as its source. Considers ethics (including but not limited to human agency and oversight, transparency, non-discrimination, accessibility, and biases and fairness) as one of the core pillars when developing or deploying AI systems.
- **Safety:** Weighs the benefits and risks before allowing third parties to process personal data. Considers the ethical consequences of AI systems throughout their life-cycle: they include both the environmental impact (environmental consequences of the production of digital devices and services) and societal impact, e.g. platformisation of work and algorithmic management that may repress workers' privacy or rights; the use of low-cost labour for labelling images to train AI systems.
- **Problem solving:** Aware that AI is a product of human intelligence and decision-making (i.e. humans choose, clean and encode the data, they design the algorithms, train the models, and curate and apply human values to the outputs) and therefore does not exist independently of humans. Aware that AI-driven speech-based technology enables the use of spoken commands that can enhance the accessibility of digital tools and devices (e.g. for those with mobility or visual limitations, limited cognition, language or learning difficulties), however, languages spoken by smaller populations are often not available, or perform worse, due to commercial prioritisation.

EduCon

Australian Framework for Generative Artificial Intelligence in Schools

- **Teaching and learning**: Generative AI tools are used to enhance teaching and learning.
- **Human and social wellbeing**: Generative AI tools are used to benefit all members of the school community.
- **Transparency:** School communities understand how generative AI tools work, how they can be used, and when and how these tools are impacting them.
- **Fairness**: Generative AI tools are used in ways that are accessible, fair and respectful.
- **Accountability**: Generative AI tools are used in ways that are open to challenge and retain human agency and accountability for decisions.
- **Privacy, security and safety**: Students and others using generative AI tools have their privacy and data protected.

EduCon

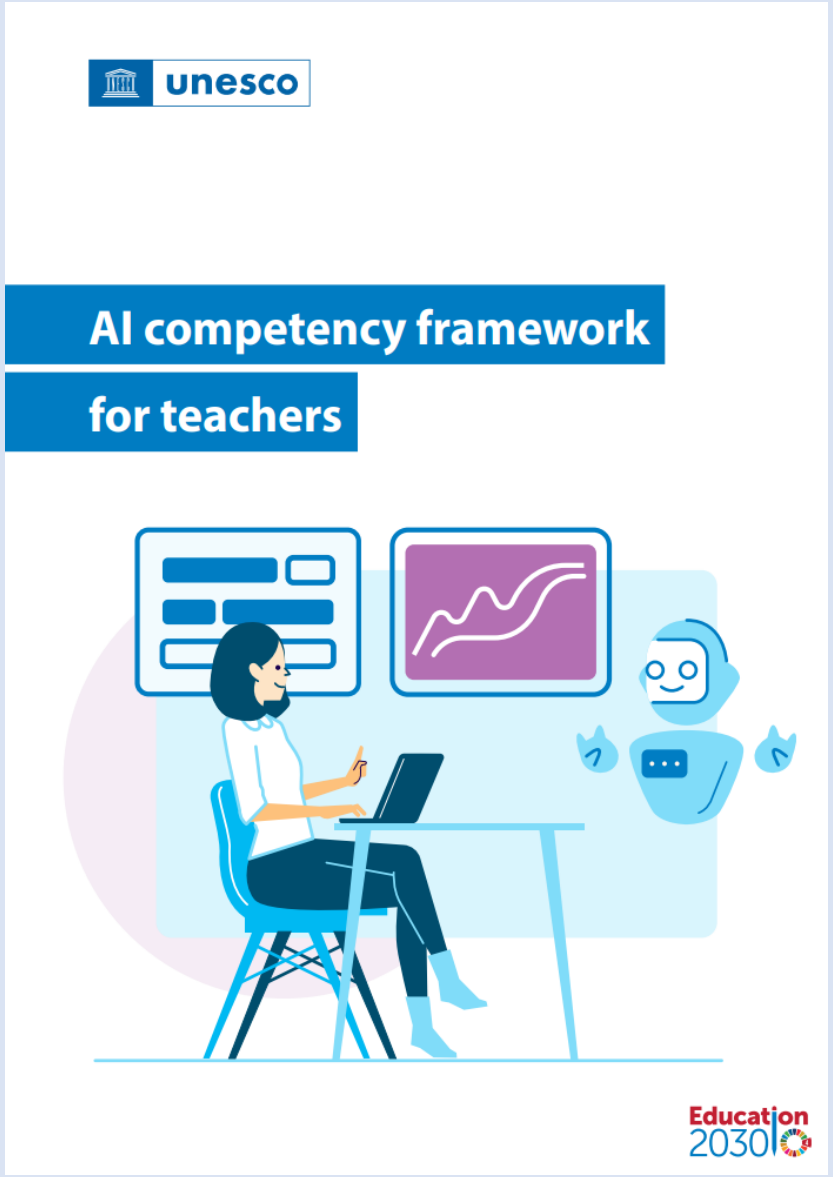
Public statement made by the Office of the Victorian Information Commissioner

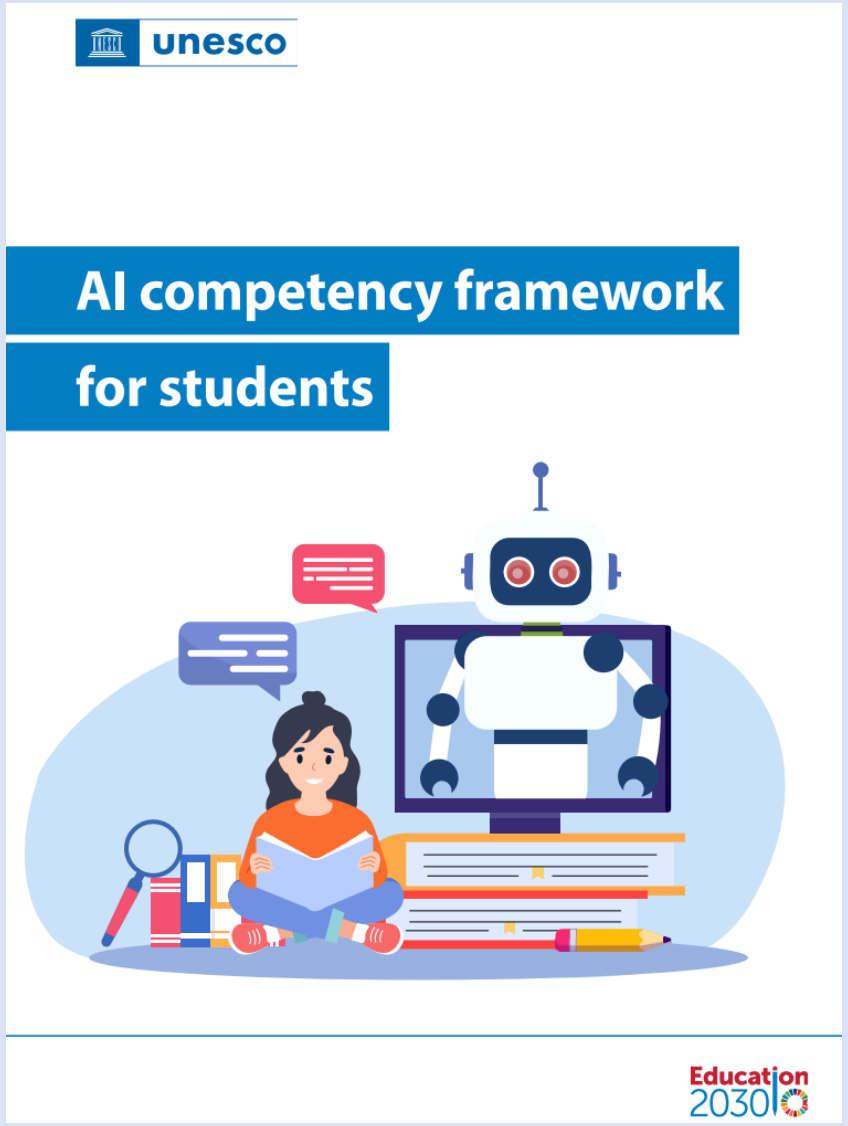
- Can personal information be used with ChatGPT (or similar GenAI tools)?
- VPS organisations must ensure staff and contracted service providers do not use personal information with ChatGPT.
- ChatGPT must not be used to formulate decisions, undertake assessments, or used for other administrative actions that may have consequences for individuals, for example, evaluations, assessments, or reviews. Doing so is a contravention of the Information Privacy Principles (IPPs), and may cause significant harm to individuals whose information is used with ChatGPT.
- If an organisation becomes aware that personal information has been used with ChatGPT it should treat the occurrence as an information security incident and notify OVIC immediately.

https://ovic.vic.gov.au/privacy/resources-for-organisations/public-statement-use-of-personal-information-with-chatgpt/

EduCon

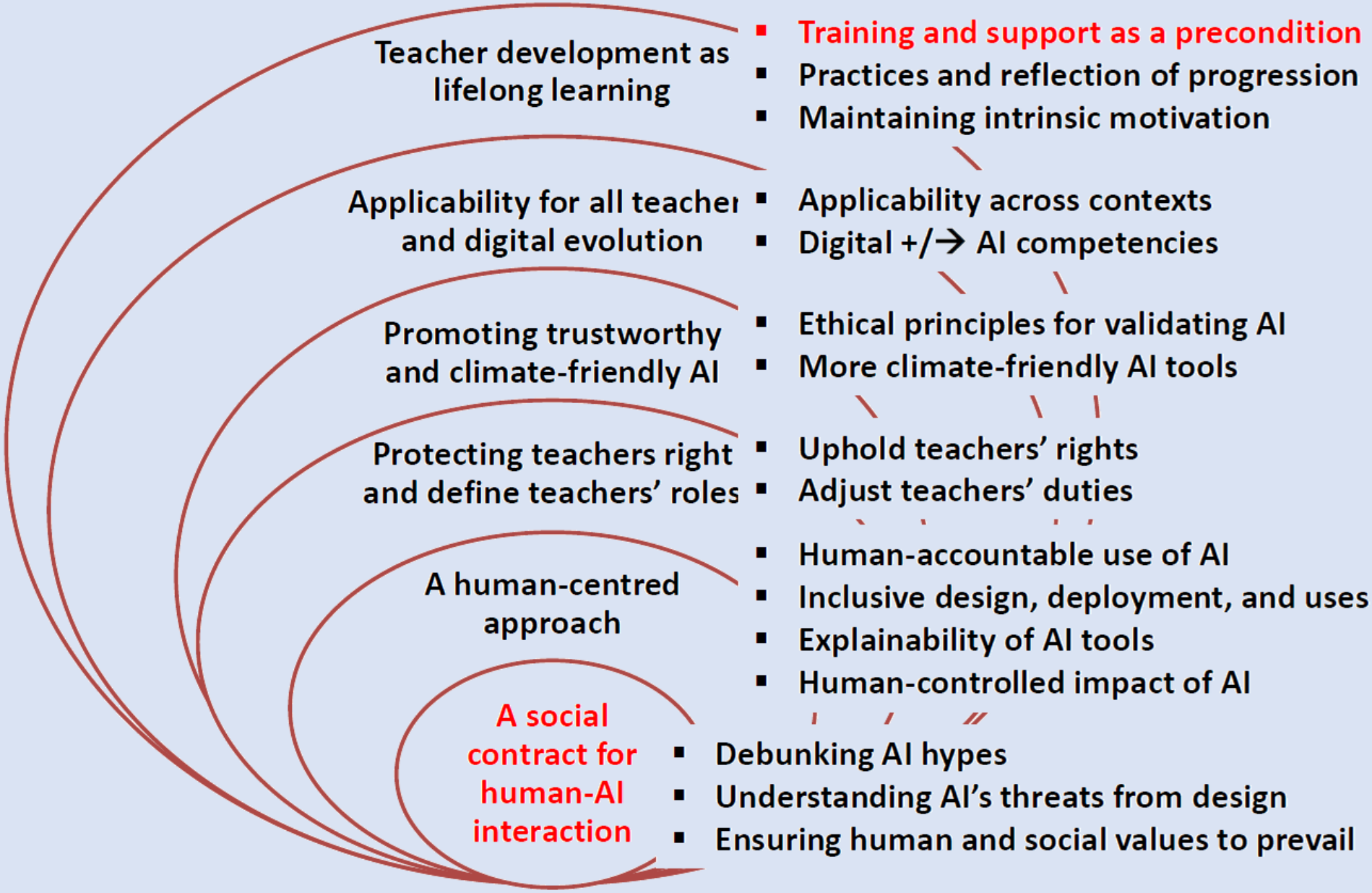
UNESCO AI competency frameworks for teachers and students





EduCon

Principles



F. Miao, UNESCO

UNESCO AI Competency Framework for Teachers

Aspects	Progression		
	Acquire	Deepen	Create
1. Human-centred mindset	Human agency	Human accountability	Social responsibility
2. Ethics of AI	Ethical principles	Safe and responsible use	Co-creating ethical rules
3. AI foundations and applications	Basic AI techniques and applications	Application skills	Creating with AI
4. AI pedagogy	AI-assisted teaching	AI–pedagogy integration	AI-enhanced pedagogical transformation
5. AI for professional development	AI enabling lifelong professional learning	AI to enhance organizational learning	AI to support professional transformation

UNESCO AI CFT aspects

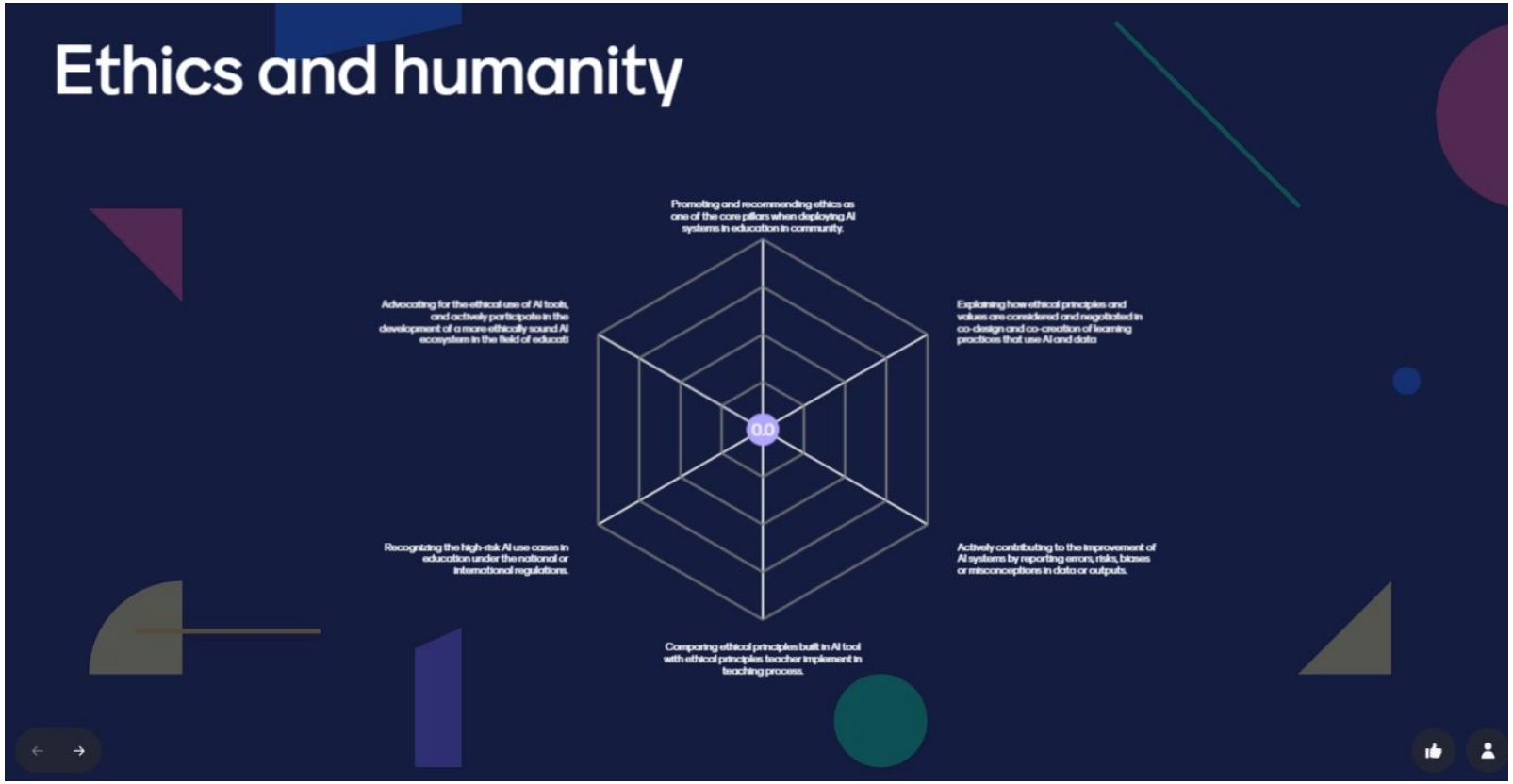
- **Human-centred mindset** defines the values and critical attitudes teachers need to develop towards human–AI interactions based on the aforementioned principles. This aspect encourages teachers to always put human rights and needs for human flourishing as the focus of AI in education. Teachers are encouraged to nurture critical methodologies to evaluate the benefits and risks of AI, while ensuring human agency and human accountability, and understanding AI's societal impact and implications for citizenship in the era of AI.
- **Ethics of AI** delineates the essential ethical values, principles, regulations, institutional laws and practical ethical rules that teachers need to understand and apply, drawn from the rapidly expanding body of knowledge on the ethics of AI and their implications for education. This aspect defines teachers' progressively deeper understanding of fundamental ethics of AI, skills to make safe and responsible use of AI, and comprehensive competencies to participate in the adaptation of ethical rules.
- **AI foundations and applications** specifies the conceptual knowledge and transferable operational skills that teachers need to understand and apply in order to support their selection, application and creative customization of AI tools to build student centred AI-assisted teaching and learning environments. Teachers are expected to gain appropriate understanding of the definition of AI, basic knowledge about how AI works, as well as about the main categories of AI technologies; the skills necessary to evaluate appropriateness and limitations of AI tools based on specific needs in specific domains and contexts; and the skills to operate validated tools for real-world tasks; progressively, it involves skills to adapt or customize AI tools to build human-centred and age-appropriate learning environments.

EduCon

UNESCO AI CFT aspects

- **AI pedagogy** proposes a set of competencies required for purposeful and effective AI–pedagogy integration, covering comprehensive competencies to validate and select appropriate AI tools and integrate them with pedagogical methods to support course preparation, teaching, learning, socialization, social caring and learning assessment. This aspect implies that teachers need to develop the ability to critically assess when and how to use AI in teaching and learning in an ethical and human-centred manner, as well as to plan and implement inclusive AI-assisted teaching and learning practices. Progressively, teachers need to enhance their capacity to critically adapt and creativity explore innovative practices in the context of advancing capabilities of emerging AI iterations.
- **AI for professional development** outlines the emerging competencies teachers need to build in order to use AI to drive their own lifelong professional learning and collaborative professional development in view of transforming their teaching practice. In response to the rapid development of AI, teachers need guidance on how to continue their professional development in educational settings characterized by growing human–AI interaction. This includes the ability to leverage AI to assess professional learning needs and nurture motivation for lifelong learning and professional collaboration. Progressively, teachers are expected to enhance their ability to adapt and create when using AI tools and data analytics to support transformative professional development.

EduCon



https://www.mentimeter.com/app/presentation/n/al8gwdksov9k6p5rxtsd5e7imsxp6is3/present

EduCon

Just because we could
doesn’t mean we should.

lidija.kralj@ucitelji.hr