How to teach responsible AI in Higher Education: challenges and opportunities

Andrea Aler Tubella[1] · Marçal Mora-Cantallops[2] · Juan Carlos Nieves[1]

Abstract

In recent years, the European Union has advanced towards responsible and sustainable Artificial Intelligence (AI) research, development and innovation. While the Ethics Guidelines for Trustworthy AI released in 2019 and the AI Act in 2021 set the starting point for a European Ethical AI, there are still several challenges to translate such advances into the public debate, education and practical learning. This paper contributes towards closing this gap by reviewing the approaches that can be found in the existing literature and by interviewing 11 experts across five countries to help define educational strategies, competencies and resources needed for the successful implementation of Trustworthy AI in Higher Education (HE) and to reach students from all disciplines. The findings are presented in the form of recommendations both for educators and policy incentives, translating the guidelines into HE teaching and practice, so that the next generation of young people can contribute to an ethical, safe and cutting-edge AI made in Europe.

Keywords Trustworthy AI · Higher Education · AI ethics · Educational strategies

Introduction

Artificial Intelligence (AI)—systems that detect and collect information from the environment and process it to solve calculations or complex problems—is an area of strategic importance for economic development and a key component of the Digital Agenda for Europe. However, the EU's High-Level Expert Group on Artificial Intelligence (AI HLEG) recognizes that "while offering great opportunities, AI systems also generate certain risks that must be managed

appropriately and proportionately" (European Commission, 2019). Therefore, it is important not only to invest in education to drive technological progress; it is also imperative that the new generations of professionals are able to shape technology in a way that respects European values. Thus, Higher Education (HE) should be tasked not only with preparing young people with advanced skills to program applications, but also to prepare all students to understand the implications of AI and to influence its ethical use.

The EU's digital strategy emphasises the need to train professionals that can "shape technology in a way that respects European values" (European Commission, 2019), explicitly acknowledging that professionals in a wide variety of fields will require knowledge of responsible AI, which has derived in great progress in defining what does Trustworthy AI mean (Floridi, 2019; Chatila et al., 2021; Kaur et al., 2022). In addition, although there is interest from HE institutions in including ethics in programs, the teaching of Responsible AI in HE still remains highly understudied and disorganised. The main goal of this paper is to assist educators in HE to introduce responsible AI into their educational programs in line with the European vision. The findings are presented in the form of recommendations both for educators and policy incentives, translating the guidelines into HE teaching and practice, so that the next generation of young

Andrea Aler Tubella, Marçal Mora-Cantallops, and Juan Carlos Nieves have contributed equally to this work.

✉ Andrea Aler Tubella
andrea.aler@umu.se

Marçal Mora-Cantallops
marcal.mora@uah.es

Juan Carlos Nieves
juan.carlos.nieves@umu.se

1 Department of Computing Science, Umeå University, 901-87 Umeå, Sweden

2 Dpto. Ciencias de la Computación, Universidad de Alcalá, Pl. de San Diego s/n, 28801 Alcalá de Henares, Madrid, Spain

people can contribute to an ethical, safe and cutting-edge AI made in Europe.

To that end, we take as a starting block the High-Level Expert Group's Guidelines on Trustworthy AI (hereon referred to as HLEG guidelines), which outline the necessary requirements for responsible and trustworthy development in the EU. It must also be noted that the present results are based on the work implemented in the context of the Erasmus+ project "Trustworthy AI",[1] a two-year project that aimed to introduce a new education framework and resources for teaching AI emphasizing the ethical and values aspects of these techniques and systems. In particular, the interviews used as base material for this work's findings were conducted as part of the first deliverable of the project, the Learning Framework for Teaching AI in HE (Aler Tubella & Nieves, 2021).

The rest of the paper is structured as follows. "Background" Sect. provides the background of this work, while the methodology used in the two phases (literature review and qualitative expert interviews) is described in "Methodology" Sect. The results are detailed in "Results" Sect., deriving into a set of recommendations for both teachers and policy-makers in "Recommendations" Sect. "Discussion and future work" Sect. discusses and compares the obtained results and recommendations to those of other works or frameworks and, finally, concludes and points to future work and limitations of the current one.

Background

In 2018 Europe decided to take the global lead in establishing a strategy for AI based on ethical principles, appropriate legal guardrails and responsible innovation. In 2019 the High-Level Expert Group on AI (HLEG) developed the Ethics Guidelines for Trustworthy AI (European Commission, 2019), defining three components which should be met throughout the system's entire life cycle. According to them, trustworthy AI should be:

1. Lawful—respecting all applicable laws and regulations;
2. Ethical—respecting ethical principles and values;
3. Robust—both from a technical perspective while considering its social environment.

The European Commission (EC) also endorsed seven principles that must be taken into consideration to assess AI systems in the White paper on AI and included them in its proposal for an AI Act; namely:

1. Human Agency and Oversight;
2. Technical Robustness and Safety;
3. Privacy and Data Governance;
4. Transparency;
5. Diversity, Non-discrimination, and Fairness;
6. Societal and Environmental Well-being; and
7. Accountability.

To operationalize Trustworthy AI, the EC introduced the HLEG trustworthy AI guidelines, which enumerate and clarify the seven requirements listed above. Transforming the HLEG guidelines into specific skills for the actors involved in AI development has been highlighted by the European Commission as a natural step in creating an "ecosystem of trust" for the flourishing of European AI (European Commission, 2020).

It is worth mentioning that the HLEG trustworthy AI guidelines are domain-independent regarding a particular AI method. New recommendations regarding Trustworthy AI are appearing in the literature; nevertheless, most of those newer recommendations identify AI principles that are either a subset or an extension of the AI principles introduced by the EC (Li et al., 2023).

One of the most significant challenges to operationalize the HLEG guidelines is to turn the trustworthy AI requirement list sectorial. This means to make the trustworthy AI requirement list suitable to the demands of a public or industrial sector, e.g., Public Safety, Health Care, Transport, Army, etc.

In terms of education, one can recognize that professionals from different sectors require also have their own interpretation of the the HLEG guidelines, but the interpretation of the guidelines requires understanding the different concepts that are introduced by the ethical guidelines.

Recently, leading Members of the European Parliament proposed to include all requirements in the AI Act to underline its main objective of ensuring that AI is developed and used in a trustworthy manner. The AI Act and the HLEG guidelines (upon which the former builds) will have a great impact on public and private parties developing, deploying, or using AI in their practices. The HLEG guidelines laid the groundwork for building, deploying, and using AI in an ethical and socio-technically robust manner, providing a framework for trustworthy AI. The AI Act further refines this framework by introducing numerous legally binding obligations for public and private sector actors, both large and small, that need to be met during the entire lifecycle of an AI system. This means that all organisations, need to start building a thorough knowledge base of trustworthy AI.

Let us mention that by reaching the vision of trustworthy AI other worldwide initiatives can also be impacted positively. For instance, some authors have analyzed the role of AI in achieving the Sustainable Development Goals (SDGs)

---

[1] https://www.trustworthyaiproject.eu/.

(Vinuesa et al., 2020). In Vinuesa et al. (2020), the authors concluded that AI has the potential to shape the delivery of all 17 SGDs, both positively and negatively. In fact, through a consensus-based expert elicitation process, Vinuesa et al. identified that "AI can enable the accomplishment of 134 targets across all the goals, but it may also inhibit 59 targets" (Vinuesa et al., 2020). Hence, one way to prevent the negative side effects of AI in achieving the Sustainable Development Goals (SDGs) is to educate professionals in the basic principles of trustworthy AI.

Methodology

Literature review

The goal of the systematic review was to analyse the relevant literature in order to answer the following questions:

1. What competences and learning objectives are identified when teaching ethical aspects in HE?
2. How are these competences taught and evaluated?

These questions were particularly selected to address the challenges of incorporating the HLEG guidelines into the classroom. Indeed, whereas the more technical aspects have more established pedagogical methods, the more abstract ethics-related content lacks concrete learning objectives and strategies.

We conducted the literature search on Scopus, to obtain results in a variety of disciplines. We used the following search terms, to be found in title, abstract or key:

```
ethics AND teaching AND "higher edu-
cation" AND ( competence OR compe-
tency OR skills ).
```

We limited our search to publications from 2015 onwards and retrieved a total of 50 publications on 09/02/2021 at 15:13. Four papers were not accessible at the time of analysis. We focused on individual research output, so excluded one book, one editorial, one extended abstract and six review articles (either paper reviews or curricula reviews). Finally, we manually excluded eight papers whose abstract did not mention anything related to teaching skills related to ethics and five papers were removed upon further reading for lack of relevance (either not focused on HE or not focused on teaching aspects related to ethics). The final output is 24 papers which we analysed (Mackenzie, 2015; Miñano et al., 2015; Trobec & Starcic, 2015; Biasutti et al., 2016; Mulot-Bausière et al., 2016; Gómez and Royo, 2015; Sánchez-Martín et al., 2017; Galanina et al., 2015; Gokdas & Torun, 2017; DeSimone, 2019; Fernandez & Martinez-Canton, 2019; Lapuzina et al., 2018; Rameli et al., 2018; Riedel and
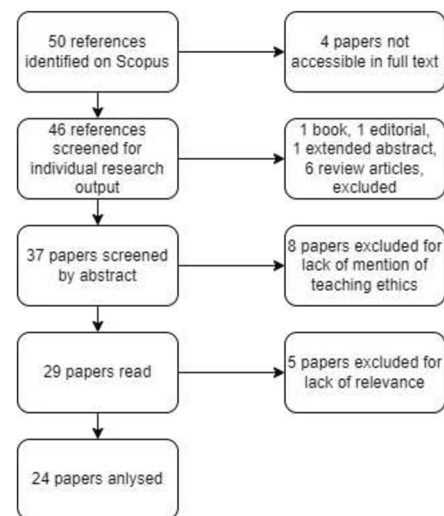

Fig. 1 Flow diagram of the selection of papers for review

Table 1 Distribution of reviewed publications by subject areas

Discipline	No. of publications
Social sciences	16
Engineering	6
Nursing	5
Medicine	4
Business, management and accounting	3
Computer science	3
Environmental science	3
Arts and humanities	2
Biochemistry, genetics and molecular biology	1
Economics, econometrics and finance	1
Energy	1
Mathematics	1

Giese, 2019; Aközer & Aközer, 2017; Oliphant & Brundin, 2019; Brown et al., 2019; Jones et al., 2020; Bates et al., 2020; Dean et al., 2020; Zamora-Polo & Sánchez-Martín, 2019; Ibáñez-Carrasco et al., 2020; Sahin & Celikkan, 2020; Noah & Aziz, 2020). Figure 1 depicts a flow diagram of the selection of paper.

The papers analysed cover a wide variety of subject areas. Based on author affiliation 17 countries are covered, as well as 12 subject areas as indexed by Scopus (Table 1).

Publications identifying specific competences are few, although many mention that explicitly identifying competences is a pressing educational need. On the other hand, most publications propose teaching methods, with a strong focus on learning with a social component of debate and participation between students. For this reason, non-traditional

teaching methods like case-studies and role-playing seem to be often proposed and studied. Much of the literature consists on exposing or evaluating how certain teaching practices were incorporated to teach ethics in specific degrees or modules. Much of the literature emphasises the importance of incorporating different dimensions of ethics into their education. In particular, professional ethics as it refers to codes of conduct is mentioned often.

A discussion of the topics emerging from the literature and how they tie in with the needs identified by educators follows in "Results" Sect.

Qualitative interviews

With the goal of exploring the state-of-the-art of Trustworthy AI in HE, we developed an interview protocol (see Appendix). The interview protocol consists of 8 sections:

1. Introduction: Introduction of the people involved in the meeting, project—background, consent issues, description of process, and follow-up steps.
2. Introduction of the Purpose of the Interview: Slide deck and agenda for the interview (approximate timeline for each section).
3. Education case: The interviewee is asked to describe an education case that will be the focus of the questioning. This case can be real or prospective, and is meant to provide contextual information.
4. General perspectives: This part is a generalised discussion of Trustworthy AI and its role within HE at large. Interviewees are asked to comment on aspects such as national or local education strategies, practices in current education, resources being used currently, and minimum incentives that should be there for promoting trustworthy AI in HE.
5. Questions on the Assessment List: This section is focused on the HLEG assessment list, which translates AI principles into an accessible and dynamic checklist intended to guide developers and deployers of AI in implementing such principles in practice. Interviewees are asked about its usefulness for education purposes, about its specific inclusion in courses, and about types of support needed to teach it.
6. Ordering of the Requirements: Participants are asked to rank the 7 Requirements in order of their application/importance in their chosen education case (with 1 being the highest). While doing so, interviewees are asked to define in their own words what each requirement entails, and to justify the ordering that they choose.
7. Questions for Specific Requirements: For the highest and lowest ranked requirement, interviewees are asked

Table 2 Interviewees' geographical location as given by affiliation (note that some interviewees have several affiliations)

Organisational affiliation	Number of interviewees
Croatia	1
Ireland	3
Netherlands	2
Spain	3
Sweden	2
EU	1

Table 3 Interviewees' profile

Position	Number of interviewees
Program coordinator	4
Professor	4
Lecturer	4
Policy maker	1

Table 4 Disciplines of the use cases selected by the interviewees

Discipline of the use case	Number of interviewees
Computer science	6
Medicine	1
Law	1
Social sciences	1
Media studies	1
Multidisciplinary	1

which aspects of it are already considered in their use case, and how. Additionally, they are asked to comment on which questions around this requirement are the most valuable ones for trustworthy AI education.
8. Closing remarks: Final wrap-up of the interview, voicing of any additional comments.

Partners from ALLAI, Universidad de Alcalá, Maynooth University and Umeå Universitet completed a training session in order to unify how the interviews were conducted. Interviewees were therefore asked the same questions in the same manner. Additionally, interviewers reported on their interviews through a standardised form, identical for each partner. All answers were contrasted in a qualitative analysis.

A total of 11 interviewees were selected for their involvement in HE, whether through governance, program management or teaching. Interviews were conducted over a period of 6 weeks. The experts, with affiliations in five different countries, brought use cases spanning medicine, law,

Table 5 Choice of most important rquirement to include in education

Requirement	Proportion of interviewees that ranked it as **most important** (%)
Human agency and oversight	18.2
Technical robustness and safety	36.4
Privacy and data governance	27.3
Transparency	9.1
Diversity, non-discrimination and inclusion	0
Societal and environmental well-being	0
Accountability	9.1

Table 6 Choice of least important requirement to include in education

Requirement	Proportion of interviewees that ranked it as **least important** (%)
Human agency and oversight	9.1
Technical robustness and safety	9.1
Privacy and data governance	0
Transparency	54.5
Diversity, non-discrimination and inclusion	0
Societal and environmental well-being	18.2
Accountability	9.1

computer science and social sciences (see Tables 2, 3 and 4). The responses from expert interviewees inform the recommendations made in this framework and shed light on the current state of Trustworthy AI in education.

Results

This section is structured around emerging themes from the qualitative interviews.

Lack of explicitness

The guidelines in their current form are valued by all interviewees for setting down clear requirements and bringing clarity to their meaning. However, their inclusion in education raises concerns: respondents raise that the length and technical nature of the documents is not suitable for all disciplines and education levels; the wording and scope of the guidelines remains abstract while using technical terminology. In addition, a perspective on how each requirement applies to different disciplines is missing: for example, how should legal scholars evaluate robustness, as opposed to

computer scientists? Furthermore, interviewees raise concerns that the translation of different technical terms may bring different perspectives depending on which language of the guidelines is being studied. A frequent point made by the experts is that different courses may touch upon only a few of the requirements, therefore not looking at the guidelines as a whole, but rather focusing on a few specific relevant requirements. Overall, several respondents note that the key aspect of the guidelines is the focus on the human behind the system and they emphasise the value of conveying to students that the responsibility and ethical obligations of AI development lies on those involved in the process.

There is significant consensus among the experts (100% of respondents) that all requirements are relevant, but that their significance and importance for a course varies depending on the topic and the area. For this reason, there was no significant agreement when they were asked to rank the seven requirements in order of importance: each education case elicited different rankings depending on the application area and topics tackled in the course (Tables 5 and 6). Despite the variation in rankings, "Transparency" stands out as being ranked the least relevant in 54.5% of the education cases. The reasons for this rating, however, are very disparate. Some experts believe that transparency is encompassed by other requirements, while others think that other requirements use more basic concepts that are easier for beginners. Additionally, some experts see the rest of the requirements as more fundamental.

In terms of which of the HLEG requirements are already being taught in current courses and programs, 60% of the experts interviewed stated that some requirements are currently included in their education case. A common thread in the expert interviews is that while different requirements are certainly covered in education, they are not explicitly related to AI or to the HLEG guidelines. Many report that topics related to trustworthiness are addressed due to their relevance, but oftentimes a deliberate effort is not made to establish an explicit relationship to the HLEG requirements. Likewise, there is no deliberate effort to include the totality of the requirements in education. When asked which requirements are currently covered in their educational case (Table 7), there is a big disparity, with Privacy and Data Governance topics being taught in 90% of the cases brought by the experts, while Societal and Environmental Well-being and Accountability are each only taught in 40% of the cases.

Need for concrete learning objectives related to RAI

The questions raised on each requirement, although specific to each, have two common threads: recognition and implementation. On one hand, there is a strong call for teaching students how to recognise whether a requirement

Table 7 Current inclusion of requirements in education

Requirement	Proportion of interviewees that state it is already included in their education case (%)
Human agency and oversight	60
Technical robustness and safety	50
Privacy and data governance	90
Transparency	60
Diversity, non-discrimination and inclusion	70
Societal and environmental well-being	40
Accountability	40

is being followed. On the other hand, many questions revolve around technical methods for trustworthy AI development, e.g., record-keeping methods, privacy preserving data collection methods, explainability methods.

Echoing this idea, when assessing competencies related to incorporating ethical and social dimensions into HE in all disciplines, results of our literature review indicate an emphasis on dual competence (Brown et al., 2019; Noah & Aziz, 2020; Trobec & Starcic, 2015; Zamora-Polo & Sánchez-Martín, 2019; Sánchez-Martín et al., 2017): developing technical competence alongside the ability to understand and act according to ethical and social expectations. Although discussion on specific learning outcomes is notably absent, three learning goals are prevalent for demonstrating mastery of social and ethical competencies:

- Ethical appreciation/sensitivity: Identifying and understanding the ethical and moral dimensions of a situation.
- Ethical analysis: Deliberating about actions, how they relate to ethical guidelines and codes of conduct, and their possible consequences.

- Ethical decision-making/Applied ethics: Selecting and implementing a course of action in response to ethical reasoning.

Some examples of how these competences are identified can be found in Table 8. These findings squarely align with syllabus analysis, where it has been found that the most common sought outcomes for teaching Tech Ethics are variations on "recognize/critique/reason" (Fiesler et al., 2020).

Thus, both the literature analysis and the expert interviews reveal the need for two different levels of expertise. The first is the call for educating on how to recognise whether a requirement is being followed and, if so, how it is being followed.. This competence corresponds to Ethical appreciation/sensitivity as identified in the literature review: understanding what a requirement means in the context of a certain application. In fact, this type of question universally applies to students as citizens, as it allows for identifying and adopting trustworthy technology. In addition, it provides an initial maturity level in terms of understanding the HLEG Requirements.

The second competence identified across requirements corresponds to technical methods for trustworthy AI development. There is consensus across interviewees about the need to teach concrete methods for explainability, traceability, data collection, impact assessment, etc. This necessity closely relates to Ethical analysis and Ethical decision-making as identified in the literature: knowledge of the available technical tools is necessary to be able to make an informed choice and implement it. Since the relevant techniques vary greatly depending on the topic and area of the course, it is particularly important to explicitly include in the curriculum which topics and methods will be addressed (Bates et al., 2020). A selection of the topics that the experts believe are necessary to teach with respect to each requirement is shown in Table 9.

Table 8 Literature quotes reflecting the three identified levels of competence

Competence	Quotes
Ethical appreciation	"When solving an ethical situation, the awareness and respect of ethical principles are not enough, the individual's role and ability to perceive the ethical dimension are also important." Trobec and Starcic (2015)
	"Using technology to assess sociopolitical, ethical, and historical issues related to nursing practice." DeSimone (2019)
	"Capacity to universalize principles of action, as a prerequisite for principled reasoning." Aközer and Aközer (2017)
Ethical analysis	"The student analyzes the pro and con viewpoints of an ethical question." DeSimone (2019)
	"reflective awareness of questions about both the "good life" and "right action" as objects of principled reasoning." Aközer and Aközer (2017)
Ethical decision-making	"Tutors also perceived technical skills and the application skills of problem-solving and critical thinking, as important." Dean et al. (2020)
	"Practice in the art of ethical decision making is the best way for ensuring that the graduates of educational leadership programs have the ethical muscle to make decisions that are moral." Jones et al. (2020)
	"Making and defending ethical decisions related to health-care issues." DeSimone (2019)

Table 9 Topics identified as relating to each requirement

Requirement	Topics
Human agency and oversight	Recognising and implementing the appropriate level of human control to guarantee human agency and oversight; Legal constraints of the application area; Human rights perspective
Technical robustness and safety	How to measure and ensure accuracy and reliability; Balancing technical robustness and ethical constraints
Privacy and data governance	How to collect and recognise quality data; Maintaining privacy; Preventing biases in the data; Legal requirements for data collection and storage (e.g., GDPR)
Transparency	Recognising transparent systems and how they differ from opaque ones; developing Explainable AI; How to document and expose the data, processes and decisions taken in the design process; Transparency of data processing: being explicit about how data is collected and what is done with it
Diversity, non-discrimination and inclusion	Recognising bias; Bias mitigation techniques
Societal and environmental well-being	Conducting and reading impact assessments
Accountability	Auditing intelligent systems; Liability and demonstration of due diligence

Lack of implementable use cases

Uniformly across interviews, experts mention that they do not use any specific resources related to Trustworthy AI. Rather, some mention the use of current topical examples, case studies, and relevant literature. A popular way to introduce Trustworthy AI concepts in the classroom is to discuss current social concerns with the applications of the technology studied in the course. In fact, 6 out of 11 interviewees believe that it would be valuable to relate the abstract requirements set up by the guidelines to more practical terms—either through real-world examples, industry participation or concrete tools to experiment with different concepts in class.

Similarly to the interview results, the literature review reveals that ethical and moral reasoning skills are often taught through student-led methods focused on encouraging reflection and debate amongst students: case studies (Lapuzina et al., 2018), role playing (Trobec & Starcic, 2015), debate (Brown et al., 2019), experiential learning (Ibáñez-Carrasco et al., 2020). This observation aligns with findings from other literature reviews, which emphasise the prevalence of games, role playing and case studies in Engineering and Computer Science Education (Hoffmann & Cross, 2021). Although wide-spread, there is however dissent in the literature, where some advocate for more formal training in, e.g., moral philosophy, in contrast to student-led activities (Aközer & Aközer, 2017).

The teaching strategies most often used to teach Trustworthy AI aspects influence the type of resources currently available. Indeed, diverse bodies have developed openly available case studies on AI ethics, such as Princeton University,[2] MIT,[3] the Markkula Center for Applied Ethics at Santa Clara University,[4] University of Washington[5] and UNESCO.[6]

When interviewees were asked about what type of resources would be useful for integrating Trustworthy AI in HE, several themes emerged. Firstly, 5 out of 11 interviewees coincided in asking for use cases. Interestingly, there was significant consensus on the type of use cases deemed necessary: they should be realistic and implementable. Indeed, using real cases brought directly from the industry that mimic situations where graduating students may find themselves in is seen as important for the usefulness of these scenarios. In contrast with the literature, where use cases are often used for reflection and debate, several interviewees suggested that use cases should be used for practical exploration, where they can implement and "play with" different solutions.

Another frequent mention is a need for material to aid in evaluation, i.e., exercises or assignments with a grading guide that can be directly used for assessing students. Indeed, several interviewees shared the difficulty of evaluating knowledge of abstract concepts. Across the literature review and the expert interviews, there is a noticeable lack of consistent methodologies for assessing soft competences such as ethical and social awareness or understanding and application of guidelines and codes of conduct. The assessment methods uncovered in the literature review mostly rely on self-assessment (Ibáñez-Carrasco et al., 2020; Mulot-Bausière et al., 2016) or experts' perception of student's knowledge without explicit grading criteria (DeSimone,

---

[2] https://aiethics.princeton.edu/case-studies/.

[3] https://computing.mit.edu/cross-cutting/social-and-ethical-responsibilities-of-computing/serc-cases-studies/.

[4] https://www.scu.edu/ethics/ethics-resources/ethics-cases/.

[5] http://techpolicylab.uw.edu/telling-stories/.

[6] https://www.unesco.org/en/artificial-intelligence/recommendation-ethics.

2019; Lapuzina et al., 2018). On the other hand, interviewees either report no explicit assessment of competences related to Trustworthy AI or include it as part of the overall assessment of programming projects.

Distance between policy and practice

Uniformly, most interviewees state that they are not aware of any specific policy strategies to include aspects of Trustworthy AI into education, either at the level of their institution or at a national level. Simultaneously, interviewees mention that the topic of Trustworthy AI is gaining importance in their organisation, and that they are actively considering how to include it in their programs. This mismatch indicates that even though Trustworthy AI is being introduced into HE, the effort is mainly driven by the educators themselves rather than by organisational or national strategies. This approach presents the risk of a mismatch in competences between programs in different HE institutions, as the introduction of Trustworthy AI into educational programs is carried out independently rather than within a coordinated strategy. This is in contrast to current strategies for AI, which highlight the need to roll out Trustworthy AI education at a national and European level. For example, the Spanish National Strategy for Artificial Intelligence states that "it is essential to ensure that students, teachers, public sector personnel, the workforce in general and society as a whole receive appropriate preparation for and training in AI, from an ethical, humanistic and gender perspective" (Ministry of Economic Affairs, n.d.). Similarly, the Swedish National Approach to AI states that "Sweden needs a strong AI component in non-technical programmes to create the conditions for broad and responsible application of the technology". Thus, despite a full acknowledgement of its importance, no coordinated effort to incorporate Trustworthy AI education is reaching educators at this time.

In terms of policy needs and incentives to boost the introduction of Trustworthy AI in HE, interviewees delivered a big variety of needs. A big point of consensus (5 out of 11 interviewees) is the need for investing in Trustworthy AI expertise so that educators are equipped to teach these topics: this can take the form of investing into multidisciplinary training or boosting the hiring of experts in Trustworthy AI aspects to participate in education. This idea aligns with interviewees mentions of lack of time to get acquainted with the topics in order to be prepared to introduce them in the classroom. When asked about risks, there was significant consensus amongst interviewees in mentioning that there is the risk of introducing Trustworthy AI in HE before institutions are able to prepare, i.e., before there is enough expertise in the topic to be able to teach it competently.

Several interviewees mention the importance of allowing for flexibility in the degree structure to allow for the inclusion of broader interdisciplinary topics. They mention that current policies strictly constrain the learning goals of different programs and leave little room for interdepartmental collaboration and interdisciplinarity. In contrast, Trustworthy AI is seen as a topic that would benefit from student's exposure to different disciplines, calling for policy incentives that will encourage interdisciplinary learning. These thoughts align with recent calls for transversal education that allows for interdisciplinarity when considering ethics in technology (Raji et al., 2021). In addition, another relevant risk mentioned is that it is important that students from all disciplines should be able to learn about Trustworthy AI. Whereas it seems that it is starting to be a focus in STEM, there were some concerns that other disciplines may not be exposed to the topic in HE. Several interviewees emphasised that aspects of Trustworthy AI are important for students not only as future professionals, but also as citizens. In this sense, they emphasised the benefits of training a generation of professionals that will possess interdisciplinary knowledge and be able to communicate with professionals from other disciplines on the terms of Trustworthy AI.

Recommendations

For the teachers

Both the literature analysis and the expert interviews reveal the need for different levels of expertise. This is justified in two ways. Firstly, social and ethical issues require different types of skills to recognise, debate, and act upon. Secondly, the interdisciplinary nature of the topics within Trustworthy AI means that students in different disciplines do not necessarily need to reach the same learning outcomes for each requirement: for some, knowledge and identification of potential issues will be necessary, whereas for others it will be necessary to master technical methods and solutions. In addition, as highlighted in "Need for concrete learning objectives related to RAI" Sect., concrete learning outcomes are notably absent in the literature and in practice, making it harder to assess concrete learning goals. Since the relevant topics and techniques vary greatly depending on the topic and area of the course, it is particularly important to explicitly include in the curriculum which topics and methods will be addressed (Bates et al., 2020) (as echoed in "Lack of explicitness" Sect.).

Thus, the recommendations for educators are as follows:

1. Explicitly include HLEG requirements in courses when relevant.

2. Bridge the gap between requirements and course content by being explicit about which requirements are being tackled in the course and how.

3. Explicitly include Trustworthy AI development methodologies in curricula (e.g., record keeping procedures, privacy-preserving data collection methods, explainability tools).

4. Set out clear learning outcomes that describe the level of proficiency expected from the student.

Recommended learning outcomes for each individual requirement are:

LO1 Appreciation: Identifying the applicability of the requirement in different contexts and its different dimensions for different stakeholders.

LO2 Analysis: Deliberating about possible implementations of the requirement, how they relate to ethical guidelines and codes of conduct, and their possible consequences.

LO3 Application: Selecting and technically implementing a solution in response to analysis in terms of the requirement.

For the policy-makers

When it comes to the effective introduction of Trustworthy AI in HE, both the incentives needed as well as the perceived risks strongly hinge on coordinated policy efforts. As reviewed in "Distance between policy and practice" Sect., current efforts to incorporate Trustworthy AI education seem to be mostly at an individual educator level. This increases the risk of unequal outcomes, as well as putting educators in the position of having to come up with Trustworthy AI curricula themselves.

Additionally, the introduction of Trustworthy AI in the classroom requires resources: time to develop curricula and learn about the topic, experts involved in education, and a multi-disciplinary perspective. All of these aspects can only be made possible through strong policy incentives that provide these resources.

Thus, we strongly encourage policy-makers to consider the following recommendations when translating the national strategies into practice:

1. Coordinate the introduction of Trustworthy AI in curricula through national education strategies, ensuring a uniform adoption.

2. Incentivise HE institutions to obtain the relevant expertise needed to teach Trustworthy AI, both by investing resources in training for educators and by hiring experts.

3. Incentivise interdisciplinary collaboration in education by valuing it in the curriculum and introducing credits for it.

Discussion and future work

Artificial intelligence is an area of strategic importance for the economic and social development of the European Union and a key component of the Digital Agenda for Europe. However, at the same time that AI systems offer immense opportunities, they create risks and may contravene our democratic or ethical principles in areas such as the agency of human beings, inclusion (or its inverse, discrimination), privacy, transparency and more. HE plays an important role in contributing to cutting-edge, safe, ethical AI. As Borenstein and Howard (2021) write, "if the technology is going to be directed in a more socially responsible way, it is time to dedicate time and attention to AI ethics education."

Both AI researchers and the organisations that employ them (mostly HE institutions) are in a unique position to shape the security landscape of the AI-enabled world. In this context (Brundage et al., 2018), highlight the importance of education, ethical statements and standards, framings, norms, and expectations, and how "educational efforts might be beneficial in highlighting the risks of malicious applications to AI researchers". In the previous sections, a list of recommendations have been formulated for both teachers and policy-makers which arise from this first attempt at translating the HLEG guidelines into HE teaching and practice.

The current findings are aligned with previous studies; for instance, Gorur et al. (2020) explored the ethics curricula of 12 Australian universities' Computer Science courses and found that, while there is wide variation in the content of ethics courses in Australian CS curricula, there tends to be a greater emphasis on professional ethics than on philosophical and macro-ethical aspects, which appear to be neglected; including the HLEG requirements in courses when relevant might alleviate such omission. It is a limitation that we do not have data on the interviews according to discipline. Thus, it would be interesting to investigate in future work which requirements are already being included in each discipline specifically.

As Aiken et al. (2000) identified years before the current focus on Ethical AI, it is essential to look beyond the student-teacher relationship, so all stakeholders should be considered as any of them might be a source of vulnerabilities or asymmetries that would increase the risk of the system. The guidelines can be the starting point for

that discussion and, thus, the appreciation, analysis and application of Ethical AI case studies are another relevant recommendation for teachers.

While many of the HLEG requirements might already be partially reflected in existing legislation, no legislation covers all of them in a comprehensive manner (Smuha, 2020), let alone in a consistent way across all European countries. In the same line, many of the ethical risks raised by the development and use by AI are context-specific which requires not only a horizontal approach but also a vertical one that holistically takes into account all the possible risks. As Smuha (2020) notes, only a very limited number of initiatives take such approach into account and even the White Paper on AI by the European Comission fails to mention the domain of Education (European Commision, 2020), which also points to the actions that are still needed in form of future policies, legislation and national (or European) education strategies.

As Dignum writes (2021), "more than multidisciplinary, future students need to be transdisciplinary—to be proficient in a variety of intellectual frameworks beyond the disciplinary perspectives". This requires a set of capabilities that are not covered by current education curricula and calls for a redesign of current studies that should start by considering (and incentivising) the relevance of interdisciplinary collaboration.

Thus, several relevant questions arise to address for the very near future. Firstly, it is important to understand how education in Trustworthy AI can be rolled out in a coordinated and consistent manner within the EU. Even though the HLEG's recommendation represent a broad European perspective, the perspective on different values will vary for each society: any effort to standardise education will need to give enough leeway for diverse perspectives.

Secondly, the question of how to "teach the teachers" arises. As many interviewees report, the competence to teach Trustworthy AI is not necessarily found within HE institutions already, not least due to the different aspects that each requirement encompasses. A strong push to develop material, courses and pedagogical guides that cover all seven requirements is therefore essential, and to the best of our knowledge there is a lack of studies on how best to introduce HE educators to these topics.

In parallel, it is important to question the idea of simply training computer science educators to provide familiarity with these topics. Indeed, a highlight between both literature and interviews is the need for interdisciplinary education. These findings echo (Raji et al., 2021), who highlight the "exclusionary default" of a purely Computer Science lens on education. For this reason, a key question to explore is how to effectively offer interdisciplinary education in Trustworthy AI. This will require understanding "the level of interaction between different disciplines and constructive alignment" (Klaassen, 2018), and in particular it will require interdisciplinary pedagogical research.

Appendix A: Interview protocol

With the goal of exploring the state-of-the-art of Trustworthy AI in HE, we developed this interview protocol. The specific goals of this protocol were to obtain expert feedback on the following topics: (1) General awareness of the Guidelines amongst stakeholders in HE (2) Inclusion of the Requirements in current educational programs (3) Current educational practices for Trustworthy AI (topics, learning outcomes, evaluation) (4). Incentives to facilitate the inclusion of Trustworthy AI topics in HE (5) Risks and opportunities Partners from ALLAI, Universidad de Alcalá, Maynooth University and Umeå Universitet followed a training session in order to unify how the interviews were conducted. Interviewees were therefore asked the same questions in the same manner, allowing to contrast answers in a qualitative analysis.

Purpose of the protocol

This protocol is meant as a guideline for interviewers. For the purposes of this document, "you" can refer to the reader (if read within a statement) or the interviewee (if read within a question).

What would be the best-case scenario? In the best-case scenario, this project helps to improve and redesign education programs in HE in the field/scope of Trustworthy AI.

As a side-effect, the interviews may help the participating entities may get a better view of the aims of the Trustworthy AI guidelines and to reflect on how these affect or are aligned with their own views and education programs. Our goal is to receive feedback about the following aspects:

1. What is the understanding of HLEG guidelines?
2. How useful is the assessment list for Trustworthy AI in education?
3. How relevant is it for HE?
4. How clear is it for HE?
5. How precise is it for HE?
6. How complete is it for HE?
7. Which issues are already covered by existing HE programs or courses?
8. Which steps would be needed to follow to introduce Trustworthy AI education in HE programs and courses?

Agenda

Subject to change, to accommodate cultural requirements:

- 09:00–09:30 Introductions, including presentation of the Trustworthy AI guidelines
- 09:30–10:00 HE program (or course) case introduction by the organisation
- 10:00–10.30 General Perspectives and Questions
- 10.30–11:00 Questions of the Assessment List
- 11:00–11:30 Ordering of the Requirements
- 11.30–12.00 Questions for Specific Areas of the Assessment List
- 12:00–12:15 Closing remarks

Introduction

Introduction of the people involved in the meeting, project—background, consent issues, description of process, and follow up steps, etc. Ensure that interview can be (voice) recorded.

Introduction of the purpose of the interview

1. Slide deck
2. Ensure that interview can be (voice) recorded. Make it clear that none other will have access to the recordings and that they will be deleted upon the completion of the project report.
3. Ensure that it is understood it is not about the performance or vision of the entity but about the suitability of the trustworthy guidelines to improve HE.
4. Make it clear that the individuals will not be noted by name anywhere. Any information, e.g., their role or location will only be used in an aggregated manner.
5. Determine whether they would like to list their AI assessment activities as part of the final report, or whether that is confidential.
6. State the agenda for the day.

Education case (education program or a course)

Discuss the HE program (or course) with them. Allow them to present the education case. Make it clear that the scenario is meant to provide contextual information. Possible topics to discuss/ask:

1. Learning outcomes.
2. Learning outcomes vs the seven requirements of the HLEG guidelines.
3. Teaching material.
4. Examination methods.

5. Heterogeneity of the students.
6. Employability of students.

General perspectives

This section is a generalised discussion of the HLEG guidelines and its assessment of AI systems during their development, deployment, and use.

1. How would you describe the current status of "trustworthy" AI in connection with HE in general? (e.g., national education strategies, practices in current education at your organisation.)
2. Can you say something about the strategy of your organisation has for AI education development? (Purpose, development, administration, recent initiatives)
3. Which of the HLEG requirements are you already teaching in your education program? Do you teach other issues related to trustworthy AI?
4. In which education cycle should trustworthy AI education need to start?
5. What resources should be available for trustworthy education in HE?
6. Are there any resources that you already use for teaching aspects related to trustworthy AI?
7. What are the minimum incentives that should be there for promoting trustworthy AI in HE?
8. What risks and opportunities do you associate with trustworthy AI in HE?
9. How could HE benefit from trustworthy AI?

Questions on the assessment list

This section aims to ask questions on the HLEG assessment list. Try to keep the discussion within reasonable time limits.

1. In which language did you read the Guidelines and the Assessment List? (Should be asked prior to the interview, but again during)
2. In overall terms, is the assessment list useful for education purposes? Why/Why not?
3. Is it beneficial or not to make it part of an actual HE course? If so, in which form? If not, why?
4. What type of support do you need to teach the Guidelines?

Ordering of the requirements

In this section, request the participants to rank the seven Requirements (Transparency, Accountability,...) in order of their application/importance (with 1 being the highest). Make it clear that the ordering is in terms of significance to their education given the education case and within the

context of this interview. You may use the printout cards and/or remind them of the seven requirements. If multiple persons/roles are taking part in the interview, you may record any notable disagreements, but only one order is permitted, i.e., the organisation's position.

1. Interpretation (their own words) of each of the seven key requirements.
2. Which TAIG requirements of the assessment list are relevant? Why/why not? In which order? Make a ranked list.
3. Why this order and why some requirements are considered less or not relevant?
4. Are there requirements in the TAIG not relevant? Why?

Questions for specific areas of the assessment list

Take the two extremes (i.e., the highest and lowest priority) from the list produced in the previous section of the interview.

1. Which aspects are already considered in their education?
2. Why is/isn't the requirement relevant for your education?
3. Evidence of addressing it:
   - (a) What are the learning outcomes related to the requirement?
   - (b) What are the evaluation methods related to the learning outcomes?
4. Do you think this requirement is clearly outlined in the Assessment List? Could you tell me how you interpret it?
5. Which questions around this requirement are the most valuable ones for trustworthy AI education?

Closing remarks

A quick wrap-up of the interview.

1. Is there anything else you would like to add/ask?
2. What was most positive from this interview?

Data avaiIbility Not applicable.

Code availbility Not applicable.

Declarations

Conflict of interest Check journal-specific guidelines for which heading to use.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Aiken, R. M., & Epstein, R. G. (2000). Ethical guidelines for AI in education: Starting a conversation. *International Journal of Artificial Intelligence in Education, 11*, 163–176.

Aközer, M., & Aközer, E. (2017). Ethics teaching in higher education for principled reasoning: A gateway for reconciling scientific practice with ethical deliberation. *Science and Engineering Ethics, 23*(3), 825–860. https://doi.org/10.1007/s11948-016-9813-y

Aler Tubella A., & Nieves, J. C. (2021). Framework for trustworthy AI education v1.0. EU Project report UMINF 23.08, Department of Computing Science, Umeå Universitet.

Bates, J., Cameron, D., Checco, A., Clough, P., Hopfgartner, F., Mazumdar, S., Sbaffi, L., Stordy, P., & de la Vega de León, A. (2020). Integrating fate/critical data studies into data science curricula: Where are we going and how do we get there? In: FAT 2020—Proceedings of the 2020 conference on fairness, accountability, and transparency, pp. 425–435 . https://doi.org/10.1145/3351095.3372832

Biasutti, M., De Baz, T., & Alshawa, H. (2016). Assessing the infusion of sustainability principles into university curricula. *Journal of Teacher Education for Sustainability, 18*(2), 21–40. https://doi.org/10.1515/jtes-2016-0012

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics, 1*(1), 61–65.

Brown, K., Connelly, S., Lovelock, B., Mainvil, L., Mather, D., Roberts, H., Skeaff, S., & Shephard, K. (2019). Do we teach our students to share and to care? *Research in Post-Compulsory Education, 24*(4), 462–481. https://doi.org/10.1080/13596748.2019.1654693

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G.C., Steinhardt, J., Flynn, C., Ó hÉigeartaigh, S., Beard, S., Belfield, H., Farquhar S., ..., Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv preprint arXiv:1802.07228

Chatila, R., Dignum, V., Fisher, M., Giannotti, F., Morik, K., Russell, S., & Yeung, K. (2021). Reflections on artificial intelligence for

humanity. In B. Braunschweig & M. Ghallab (Eds.), *Trustworthy AI* (pp. 13–39). Springer.

Dean, B. A., Perkiss, S., Simic Misic, M., & Luzia, K. (2020). Transforming accounting curricula to enhance integrative learning. *Accounting & Finance, 60*(3), 2301–2338. https://doi.org/10.1111/acfi.12363

DeSimone, B. B. (2019). Curriculum redesign to build the moral courage values of accelerated bachelor's degree nursing students. *SAGE Open Nursing.* https://doi.org/10.1177/2377960819827086

Dignum, V. (2021). The role and challenges of education for responsible AI. *London Review of Education, 19*(1), 1–11.

European Commission. (2019). *Directorate-General for communications networks content and technology: Ethics guidelines for trustworthy AI.* Publications Office.

European Commission. (2020). White paper on artificial intelligence: A European approach to excellence and trust. Com (2020) 65 Final.

Fernandez, F.J.I., & Martinez-Canton, A. (2019). Ethics and transversal citizenship in the teaching of science and engineering. In: 2019 9th IEEE integrated STEM education conference, ISEC 2019, pp. 104–110 . https://doi.org/10.1109/ISECon.2019.8881965

Fiesler, C., Garrett, N., & Beard, N. (2020). What do we teach when we teach tech ethics? A syllabi analysis. *Conference: SIGCSE '20: The 51st ACM technical symposium on computer science education.* . https://doi.org/10.1145/3328778.3366825

Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence, 1*(6), 261–262.

Galanina, E., Dulzon, A., & Schwab, A. (2015). Forming engineers' sociocultural competence: Engineering ethics at tomsk polytechnic university. In: IOP Conference Series: Materials Science and Engineering, vol. 93. https://doi.org/10.1088/1757-899X/93/1/012078

Gokdas, I., & Torun, F. (2017). Examining the impact of instructional technology and material design courses on technopedagogical education competency acquisition according to different variables. *Educational Sciences Theory & Practice, 17*(5), 1733–1758. https://doi.org/10.12738/estp.2017.5.0322

Gómez, V., & Royo, P. (2015). Ethical self-discovery and deliberation: Towards a teaching model of ethics under the competences model—Autodescubrimiento ético y deliberación: Hacia un modelo de enseñanza de la ética en el modelo por competencias. *Estudios Pedagogicos, 41*(2), 345–358. https://doi.org/10.4067/s0718-07052015000200020

Gorur, R., Hoon, L., & Kowal, E.(2020). Computer science ethics education in Australia—a work in progress. In: 2020 IEEE international conference on teaching, assessment, and learning for engineering (TALE) (pp. 945–947). IEEE.

Hoffmann, A.L., & Cross, K.A.(2021). Teaching data ethics: Foundations and possibilities from engineering and computer science ethics education. Technical report . Retrieved from http://hdl.handle.net/1773/46921

Ibáñez-Carrasco, F., Worthington, C., Rourke, S., & Hastings, C. (2020). Universities without walls: A blended delivery approach to training the next generation of HIV researchers in Canada. *International Journal of Environmental Research and Public Health, 17*(12), 1–12. https://doi.org/10.3390/ijerph17124265

Jones, K. D., Ransom, H., & Chambers, C. R. (2020). Teaching ethics in educational leadership using the values-issues-action (VIA) model. *Journal of Research on Leadership Education, 15*(2), 150–163. https://doi.org/10.1177/1942775119838297

Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: A review. *ACM Computing Surveys.* https://doi.org/10.1145/3491209

Klaassen, R. G. (2018). Interdisciplinary education: A case study. *European Journal of Engineering Education.* https://doi.org/10.1080/03043797.2018.1442417

Lapuzina, O., Romanov, Y., & Lisachuk, L. (2018). Professional ethics as an important part of engineer training in technical higher education

institutions. *New Educational Review, 54*(4), 110–121. https://doi.org/10.15804/tner.2018.54.4.09

Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys, 55*(9), 1–46.

Mackenzie, M. L. (2015). Educating ethical leaders for the information society: Adopting babies from business. *Advances in Librarianship, 39*, 47–79. https://doi.org/10.1108/S0065-283020150000039010

Miñano, R., Aller, C. F., Anguera, A., & Portillo, E. (2015). Introducing ethical, social and environmental issues in ICT engineering degrees. *Journal of Technology and Science Education, 5*(4), 272–285. https://doi.org/10.3926/jotse.203

Ministry of Economic Affairs and Digital Transformation (MINECO;Ministerio de Asuntos Económicos y Transformación Digital): National AI Strategy. https://oecd.ai/en/wonk/documents/spain-national-ai-strategy-2020

Mulot-Bausière, M., Gallé-Gaudin, C., Montaz, L., Burucoa, B., Mallet, D., & Denis-Delpierre, N. (2016). Formation des internes en médecine de la douleur et médecine palliative : bilan et suggestions des étudiants. *Médecine Palliative : Soins de Support—Accompagnement—Éthique, 15*(3), 143–150. https://doi.org/10.1016/j.medpal.2015.12.002

Noah, J. B., & Aziz, A. B. A. (2020). A case study on the development of soft skills among TESL graduates in a University. *Universal Journal of Educational Research, 8*(10), 4610–4617. https://doi.org/10.13189/ujer.2020.081029

Oliphant, T., & Brundin, M. R. (2019). Conflicting values: An exploration of the tensions between learning analytics and academic librarianship. *Library Trends, 68*(1), 5–23. https://doi.org/10.1353/lib.2019.0028

Raji, I.D., Scheuerman, M.K., Amironesei, R.(2021). "You can't sit with us": Exclusionary pedagogy in AI ethics education. In: FAccT 2021—Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 515–525). ACM. https://doi.org/10.1145/3442188.3445914.

Rameli, M. R. M., Bunyamin, M. A. H., Siang, T. J., Hassan, Z., Mokhtar, M., Ahmad, J., & Jambari, H. (2018). Item analysis on the effects of study visit programme in cultivating students' soft skills: A case study. *International Journal of Engineering and Technology (UAE), 7*(2), 117–120. https://doi.org/10.14419/ijet.v7i2.10.10968

Riedel, A., & Giese, C. (2019). Development of ethical competence in (future) nursing programs-reification and steps as a foundation for curricular development—Ethikkompetenzentwicklung in der (zukünftigen) pflegeberuflichen Qualifizierung—Konkretion und Stufung als Grundlegung für. *Ethik in der Medizin, 31*(1), 61–79. https://doi.org/10.1007/s00481-018-00515-0

Sahin, Y. G., & Celikkan, U. (2020). Information technology asymmetry and gaps between higher education institutions and industry. *Journal of Information Technology Education: Research, 19*, 339–365. https://doi.org/10.28945/4553

Sánchez-Martín, J., Zamora-Polo, F., Moreno-Losada, J., & Parejo-Ayuso, J. P. (2017). Innovative education tools for developing ethical skills in university science lessons: The case of the moral cross dilemma. *Ramon Llull Journal of Applied Ethics, 8*, 225–245.

Smuha, N. A. (2020). Trustworthy artificial intelligence in education: Pitfalls and pathways. Retrieved from SSRN 3742421

Trobec, I., & Starcic, A. I. (2015). Developing nursing ethical competences online versus in the traditional classroom. *Nursing Ethics, 22*(3), 352–366. https://doi.org/10.1177/0969733014533241

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications, 11*(1), 1–10.

Zamora-Polo, F., & Sánchez-Martín, J. (2019). Teaching for a better world: Sustainability and sustainable development goals in the

construction of a change-maker university. *Sustainability (Switzer-land)*. https://doi.org/10.3390/su11154224