12-2023

Ethical considerations for artificial intelligence in educational assessments

LIM MING SOON TRISTAN
Singapore Management University, tristanl.2021@engd.smu.edu.sg

GOTTIPATI Swapna
Singapore Management University, SWAPNAG@smu.edu.sg

Michelle L. F. CHEONG
Singapore Management University, michcheong@smu.edu.sg

This is a preprint version. The final version has been accepted as a book chapter in the book ***Creative AI Tools and Ethical Implications in Teaching and Learning*** published by IGI Global.

To cite the article:

Lim, T., Gottipati, S., & Cheong, M. L. (2023). Ethical Considerations for Artificial Intelligence in Educational Assessments. In *Creative AI Tools and Ethical Implications in Teaching and Learning* (pp. 32-79). IGI Global. DOI: https://doi.org/10.4018/979-8-3693-0205-7.ch003

INTRODUCTION

Artificial intelligence in education (AIED) is the machine mimicry of human-like consciousness and behavior to achieve educational goals, through the use of technology that allows digital systems to perform tasks commonly associated with intelligent beings.

Of the three pillars of education, assessment exists as an important component, alongside pedagogy and curriculum (Hill and Barber, 2014). Within the AIED domain, Chaudhry and Kazim (2022) scoured the landscape and concluded that assessment is one of the four key sub-domains in AIED, alongside learning personalization, automated learning systems, and intelligent learning environments. In an educational context, assessment refers to '*any appraisal (or judgment or evaluation)… of work or performance*' (Sadler, 1989). The infusion of artificial intelligence (AI) in assessments has grown significantly in recent years. Research on assessments related to digital education in the higher education landscape showed that AI and adaptive learning technologies have tripled between 2011 to 2021 and is likely to surpass immersive learning technologies as a prime research area in the near future (Lim, Gottipati and Cheong, 2022, p. 5). Among stakeholders, there is a consensus positive view that "*AI would provide a fairer, richer assessment system that would evaluate students across a longer period of time and from an evidence-based, value-added perspective*" (Luckin, 2017).

Infusion of AI in assessments also brings along its own set of concerns. AI implementation comes with technical and operational issues relating to system implementation. Arguably, these challenges have relatively lesser grey areas to contend with, than the complication of navigating the parameters and boundaries of ethics. Evaluators, as practitioners of assessments, will need to acknowledge, respect, and uphold ethical principles that may plague the implementation of an AI-based assessment.

The objective of this chapter is to examine the landscape of AI-related ethical issues for educational assessments, through the lens of a systematic literature mapping approach. A systematic literature mapping study is a study concerned with the mapping and structuring of a topical research area, the identification of gaps in knowledge, and the examination of possible research topics (Petersen, Vakkalanka and Kuzniarz, 2015).

This chapter investigates the following research questions:

- *RQ1:* Where do the studies that discuss ethical issues relating to AI-based assessments arise from?

  This question looks at where the studies discussing ethical issues relating to AI-based assessments arise from, studies patterns arising from exploratory data analysis, and seeks to provide recommendations (if any).

- *RQ2:* What are the main AI use cases relating to assessments?

  This question looks at AI applications in different areas of assessments, and how dominantly each AI application areas are featured in related studies.

- *RQ3*: What are the main ethical issues arising from the AI implementations relating to assessments?

  This question looks at the key ethical principles related to AI applications in assessments, and how dominantly each ethical principles are featured in related studies.

- *RQ4:* What are the key themes of the systematic literature map?

  This question looks to identify key themes of the systematic literature map, and draw up a framework to visualize and generalize the key themes for researchers and practitioners.

Through a systematic meta-analysis of existing literature, this chapter helps: (i) understand and consolidate knowledge regarding what was previously explored relating to AI-based assessment methods and their interconnected ethical issues, (ii) provide an integrated inquiry into the association of the ethical problems faced, and (iii) identify potential future research topics in the field.

The study identified five key research archetypical themes, with presence across the system layers of cognitive, information and physical domains of an AI-based assessment pipeline, namely: (i) AI system design and check for assessment purposes; (ii) AI-based assessment construction and rollout; (iii) data stewardship and surveillance; (iv) administration of assessments using AI systems; and (v) AI-facilitated assessment grading and evaluation. Ten AI ethics principles epitomize the key ethics considerations across each of the five research themes; each manifesting varying levels of importance.

We hope the findings of this chapter can provide researchers and practitioners the insights into the application methods of AI in assessments, especially in terms of their intertwined ethical challenges and how these challenges may be addressed, for follow up studies.

The remainder of the chapter is organized as follows: (i) the *Background and Literature Review* section introduces the background of AIED, in relation to ethics and assessments, supported by a survey of the state-of-the-art; (ii) the *Methodology* section discusses the systematic literature mapping approaches undertaken, explains the machine learning methods utilized, discusses research validity and repeatability issues, and highlights limitations to the research; (iii) the *Findings* section presents the tables and graphic visualizations from charting, coding, topic modelling, and network analyses, and provides in-depth analyses of the data.; (iv) the *Discussion* section aims to generalize the results into key themes that can be used by practitioners and researchers, and discuss practical and theoretical implications of the findings; and last but not least, (v) the *Conclusion* section summarizes the key findings, impact, and closes with proposed future work that can be studied by practitioners and researchers.

BACKGROUND

AIED

Depending on the context in educational technology-related papers, the term 'AI' in education is commonly used broadly and interchangeably with the term 'adaptive learning technology', and to some extents 'learning analytics', 'educational data mining', 'educational data science', 'teaching analytics', 'data-driven decision-making in education', and 'big data in education' (Romero and Ventura, 2020).

It is useful to note that 'intelligence' in AI exists as a continuum. Chassang et al. (2021) describes how AI can be mapped into Bloom's taxonomy of learning, with 'Crystallized Intelligence' describing lower order thinking skills and 'Fluid Intelligence' describing higher order thinking skills. The former includes mainly supervised learning (or target-based prediction), with "*encoding capacity, middle-long term memory and ability to access memorized data in a logical way*", while the latter includes a higher level of intelligence abstraction, with "*the ability to solve new problems, use logics in new situations and identify patterns without necessarily having the prior experience of similar information or problems*".

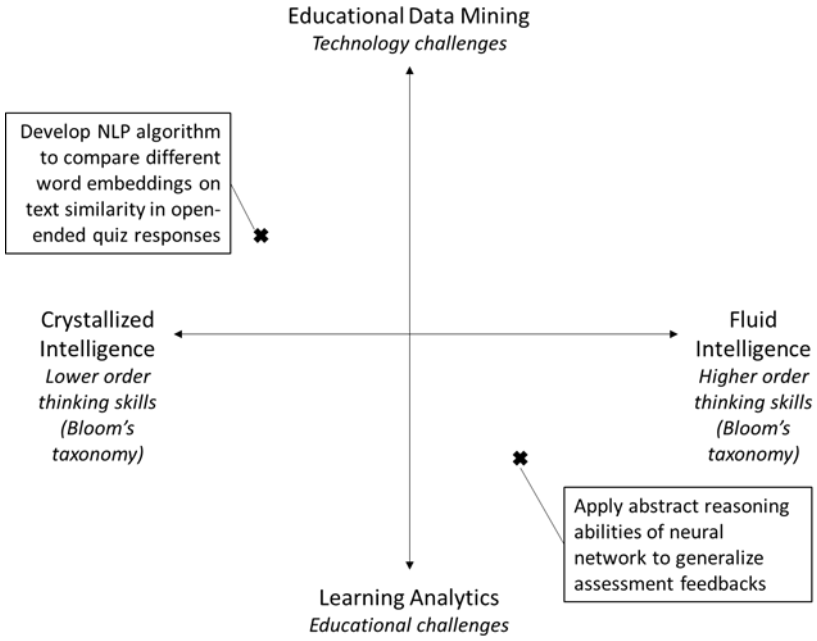

Figure 1. Relationship between educational data mining and learning analytics within AIED

AIED is a superset that encompasses terms such as educational data mining or learning analytics, depending on its specific use cases (Fig. 1). Educational data mining is relatively more focused on the technological challenges of developing and applying data mining techniques in education. For instance, the use of educational data mining that involves developing a natural language processing algorithm to compare different word embeddings on text similarity in open-ended quiz responses is a use case of AI described as Crystallized Intelligence. On the other hand, learning analytics is more focused on the educational challenges of data-driven decision-making, through the use of predictive models. For instance, the use of learning analytics applying the abstract reasoning abilities of neural network to generalize assessment feedbacks will be a use case of AI described as Fluid Intelligence. There are many other use cases of AI, including visual data analytics for AI-integrated communication dashboards and recommender systems for formative assessment questions etc.

From a technology point of view, under the sub-fields of AI are machine learning, and in turn, deep learning and reinforcement learning. Machine learning is a field of inquiry that seeks to understand and create methods to leverage on data and learn to make decisions. Deep learning

leverages on a class of machine learning methods (specifically neural networks) to identify representation elements of a dataset, so as to learn features and perform tasks. Reinforcement learning seeks to utilize machine learning methods for intelligent agents to take actions in an environment to optimize some notion of reward.

A multitude of learning techniques and algorithms exist under each of these sub-fields. For instance, collaborative analytics refers to machine learning procedures performed to measure metrics tied to interdependent student relationships to predict collaboration dynamics in group assessments. A use case in Martinez-Maldonado et al. (2021) used natural language processing techniques, including a latent semantic analysis algorithm, to study the progression of collaborative critical thinking skills using online forum data. This computer-supported collaborative learning research, was underpinned by theoretical model of communities of enquiry, positing that meaningful learning on online forums occur when there exist high levels of social, cognitive and teaching presence.

Generative AI models, which are designed to create new content or simulations, are increasingly being explored in educational settings. These models can generate personalized learning materials, simulate complex scenarios for assessment, and provide adaptive feedback, bridging the gap between traditional assessments and dynamic, individualized learning experiences.

AIED has grown tremendously in research intensity in recent years. It was observed that there was a significant increase in the number of research papers cited in Google Scholar; research publications increased from 1,739 papers in the decade of 1990 to 1999, to 22,060 papers in the most recent decade ending this year, representing more than twelve folds of increase.

2.2 AIED and ethics

At present, although AI has yet to achieve comparability mimicking human levels of consciousness, it is still of urgent and paramount importance to consider ethical issues in AI applications, including AIED.

Aside from concerns regarding infringement of relevant laws and AI crimes (Sibai, 2020), ethical threats may exist, for instance, in the forms of systemic inequality and discrimination against marginalized learner groups in AI-driven assessments. Chaudhry and Kazim (2022) emphasizes that "*risks of AI going wrong have increased significantly for all stakeholders including, ed-tech companies, schools, teachers and learners…. a lot more work needs to be done on ethical AI in learning contexts to mitigate these risks*." Prioritizing ethics is crucial to ensure the wellbeing of students, educators and other stakeholders involved in AIED.

Ethics is branch of philosophy. At its core are the concepts of "good" or "bad", and "right" or "wrong". Ethics is closely tied to the study of values (i.e., axiology) and the study of taste and beauty (i.e., aesthetics). Ethics can be subdivided into three core research areas, which can be further divided into a multitude of sub-branches. These three core research areas are, namely:

- *Metaethics:* Study of the nature (i.e., moral ontology), meaning (i.e., moral semantics), and the scope and knowledge to defend or support (i.e., moral epistemology) moral judgments.

- *Moral ontology*: This investigates the nature of moral judgments. For instance, applying moral relativism in moral ontology, we may ask the question: "*Is the right or wrong of plagiarism necessarily contextualizable to societal conventions?*"

- *Moral semantics*: This evaluates the meaning and implications to the meaning of moral judgment. For instance, applying ethical naturalism in moral semantics, it may hold that the cognitivist ethical proposition of "*ensuring fairness in assessments is an ethically good act*" may be reducible and supervene into the natural property of "*maximizing happiness*", as embraced by utilitarianism.

- *Moral epistemology*: This studies and justifies moral knowledge. For instance, applying ethical rationalism in moral epistemology, the ethical principle of cheating is tied to the moral truths of justice and fairness, which are known by reasoning alone *a priori*.

- *Normative ethics:* Study of the moral rules and standards that guide how individuals, institutions and societies should behave in a moral sense.

  - *Virtue ethics*: This emphasizes the inherent disposition of an individual, and not specific actions. 'Good' in this context is the development of practical wisdom, and the flourishing of individual character and wellbeing. As such, morality becomes a holistic personal development process. Virtual ethics argues that virtuous individuals through good motivations can make good moral choices. However, the downsides are that individuals may not agree on what is good, and the existence of value plurality is reduced.

    Here, it can be argued that ethics training for stakeholders implementing assessments in AIED achieves *phronesis* (i.e., acquiring practical wisdom to make 'good' decisions), and is sufficient in itself as an ethics stewardship measure.

  - *Deontological ethics*: This emphasizes on an individual's rights and duties, including the presence of natural, absolute rights (i.e., natural rights theory), the presence of human rationality and inviolable moral laws (i.e., Kantian categorical imperative), and the morality of good actors arising from unbiasedness behind a veil of ignorance (i.e., contractualism). 'Good' in this context is the fulfilment and discharge of moral duties. As such, morality is focused on intention and obligation. Deontology argues that clear moral intuition and boundaries exist, even in cross cultural settings, although downsides are the lack of flexibility and possibility of conflict between human rights and moral duties.

    Here, ethics should be viewed from the lens of human rights, rationality and unbiasedness. For instance, the right to privacy may be viewed as an inviolable moral standard.

  - *Consequentialism*: This emphasizes that the outcome of an action defines the morality of an action. Utilitarianism promotes actions that maximize happiness for the greatest number of people. Intellectualism promotes actions that

encourage and cultivate knowledge. Situational ethics promotes *moral particularism,* which focuses on contextualizing actions that seek to engender love. 'Good' in this context is the actions that promote 'ideal' outcomes. As such, morality becomes results focused. Consequentialism argues a practical approach that is multi-perspective and objective, although the downsides may be an over-endorsement of value pluralism and the presence of adverse intended motivation behind an action that may use consequentialism to inappropriately justify their course of action.

Here, in the case of AIED, it can be argued that the impact to the relevant stakeholders should be measured, using appropriate metrics, as an ethics stewardship measure.

- *Applied ethics:* Study of the practical application of philosophical tools to examine and provide solutions to real world morality issues.

  This can be applied on scopes of digital ethics, defined as the "*attempt to guide human conduct in the design and use of digital technology*", and in narrower terms, AI ethics, defined as the "*attempt to guide human conduct in the design and use of artificial automata or artificial machines, or computers in particular, by rationally formulating and following principles or rules that reflect basic individual and social commitments and our leading ideals and values*" (Hanna and Kazim, 2021). On the subject of definition, it is useful to highlight the difference between AI ethics and ethical AI. Siau and Wang (2020) clarifies the former as "*principles, rules, guidelines, policies and regulations related to AI*", and the latter as "*AI that performs and behaves ethically*". The former relates to the behavior of humans, whereas the latter relates to the behavior of AI systems.

This chapter does not seek to argue the meta-ethics and normative ethics tied to assessments applying AIED. In this chapter, we take a more practical approach by considering fundamental ethical principles (e.g., fairness and cheating) that inform the design, regulation and the use of AIED in assessments. These are ethics principles which provide concrete property instantiations of applied ethics, as opposed to abstract moral universals (Stringer, 2018). The chapter makes explicit these principles by describing applied instances of these principles found in existing peer-reviewed literature. In addition, the chapter cites practical solutions and mitigation measures that can be used to uphold these principles. It should be highlighted that while this chapter seeks to provide a generalizable approach to the consideration of AI ethics in assessments, the application in specific domains (e.g., medicine and healthcare) may vary in breadth (e.g., safeguard human safety during assessments in medical field training) and are outside the scope of this research.

From the real-world practical application standpoint, we note that there is a tradeoff between the agenda of advancing the AI technology, and the governance and stewardship of the use of AI in an ethical manner. Many professional and governmental bodies have pushed for responsible AI governance and stewardship. Siau and Wang (2020) identified eight institutions that have drawn up such ethics guidelines to facilitate the adoption, development and embracing of AI, including professional bodies such as the Institute of Electrical and Electronics Engineers (IEEE, 2019), and government bodies, such as the Australian government's Department of Industry, Science and Resources (Australian Government, 2019). Specifically, on the subject of AIED, Nguyen et al. (2022) studied five such relevant guidelines,

including UNESCO (2021)'s adoption on AI ethics guidelines with applications on AIED, and European Parliament (2021)'s report on AIED. These said, Siau and Wang (2020) notes that companies and institutions are presently more heavily weighted towards growing AI capabilities, with lower focus on ethical considerations.

Hinderance to the lack of focus on AI ethics in practice could stem from several reasons. Firstly, among nascent studies that investigate this tradeoff, Bessen, Impink, and Seamans (2022) studied the cost of integrating ethics in AI development, from a data management perspective. We note that further AI ethics research on cost-benefit analyses can be useful to help balance this tradeoff, and advance AI ethics governance and stewardship. Secondly, while the present AI ethics guidelines, especially the ones related to AIED studied by Nguyen et al. (2022), seek to address AI ethics issues, we note the lack of specificity on AIED applications, for instance, in assessments. Stahl, Timmermans, and Mittelstadt (2016) shares how ethics discourses should be "*focused on particular technologies to have practical importance*." In turn, Whittlestone et al. (2019) argues that these guidelines are "*not specific enough to be action guiding.*" This lack of idiosyncrasy and relevance can deter actionable applications.

To our best knowledge, a systematic literature mapping on ethical dimensions of the application of AI in assessments is lacking. This chapter aims to address the latter to enhance real world adoption in the sub-domain of assessments within AIED.

2.3 AIED and assessment

There are different applications of AI in assessment practices. Sánchez-Prieto et al. (2020) presents a systematic literature review on AI-driven assessments, and subdivided assessments into three themes, namely:

- *Assessment of student behavior:* This includes the contextualization of assessment delivery, and the prediction of assessment outcomes.
- *Assessment of student sentiment:* This includes the personalization of feedbacks, and the analysis of socio-emotional elements.
- *Assessment of student achievement:* This includes the automation of grading, and the categorization or profiling of students using data from assessment performance.

In a similar study, González-Calatayud, Prendes-Espinosa and Roig-Vila (2021) identified thematic uses of AI in assessments, mainly in individual or group adaptive formative assessments, automated grading and personalized feedbacks.

There exist different AI-driven assessment types, such as individual or group cognitive assessments and socio-emotional assessments. In addition, there exist different AI use cases across the assessment development and delivery pipeline, including but not limited to, assessment construction, curation and delivery, proctoring, grading, learning intervention and assistance, and feedback (assessment pipeline). From a technology system perspective, the AI system development pipeline includes "*the decision to start collecting data till the point when the machine learning model is deployed in production*" (Chaudhry and Kazim, 2022).

Different assessment types and areas of the assessment pipeline can be associated with different AI ethical concerns. For instance, in relation to assessment type, a socio-emotional assessment may be constructed in a manner that performs a semestral long behavioral tracking surveillance

on learners, resulting in privacy infringement and anxiety disorders among students. In another example, in relation to the assessment pipeline, specifically regarding the design of an AI assessment system, data and/or modelling deficiencies may perpetuate stigmatization of minority group students, resulting in negative learning and psychological impact.

In this chapter, we seek to discuss the ethical dimensions of the application of AI in assessments across the assessment development and delivery pipeline, taking into account different assessment types.

In closing, in this literature review, the discourse on *AIED* provides an overview of the field of inquiry, the discourse on *AIED and ethics* provides the context and scope of ethics considerations within AIED, and the discourse on AIED and assessment looks at the applications of AI in assessment practices. The next section discusses the details of the methodology applied in this systematic literature mapping approach.

3.0 Methodology

In this study, we apply the systematic literature mapping approach. The study was conducted using the research methodology in Kabudi, Pappas and Olsen (2021), building upon the guidelines as proposed by Petersen, Vakkalanka and Kuzniarz (2015). We apply the methodology undertaken by both studies as follows, namely: (i) search and selection, (ii) data extraction, (iii) classification and analysis, and (iv) evaluation of validity.

PRISMA approach, or the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses approach*, was employed as a guideline to conduct the search and selection phase (Moher et al., 2009). In accordance with the recommended methodology as part of the PRISMA-P checklist, details including the eligibility criteria, sources of information, search protocol, research records, data items and synthesis of data are described in the following sub-sections.

NVivo11, EndNote X9 and Excel spreadsheets were used for information organization. Further information extraction, data visualization, and machine learning tools and techniques are described in the following sub-sections.

3.1 Search and Selection

As AIED researchers stem from a variety of fields publishing across a wide range of publications, literature search was conducted using Scopus, an interdisciplinary rigorously curated database covering the widest range of disciplines (240 disciplines) relative to similar citation databases, with contents including over 87 million publication items, 1.8 billion cited references, 17 million author profiles, 94,000 affiliation sources and 7,000 publishers. On average, each paper indexed on Scopus has 10% to 15% more citations than similar databases (Elsevier, 2022), which implies a more extensive systematic literature mapping analysis.

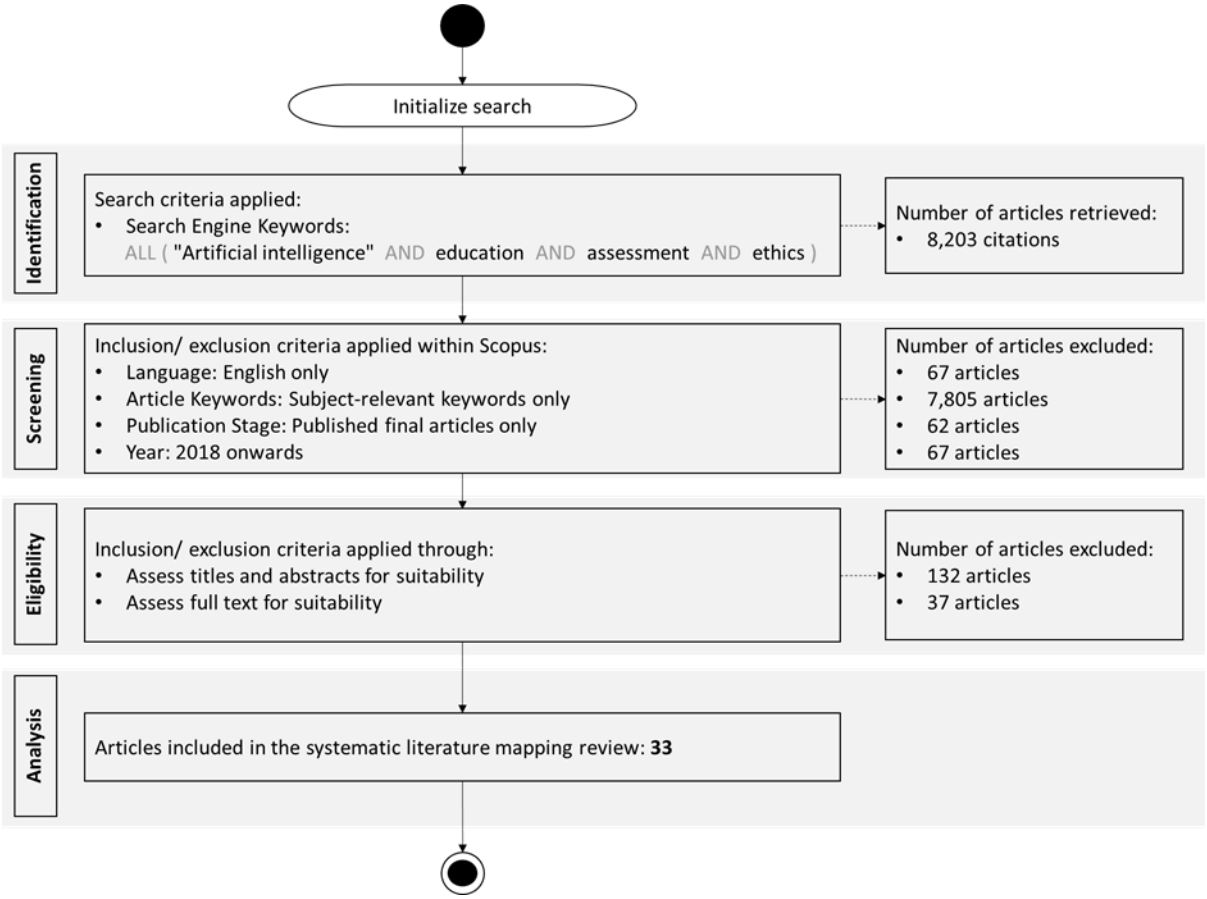

Figure 2. PRISMA - The systematic mapping process

The first stage of PRISMA, or the *identification* stage, identifies the possible papers to be considered using the Scopus search engine. The search entry was as follows: *ALL ("Artificial intelligence" AND education AND assessment AND ethics)*. This stage identified a corpus of 8,203 papers.

The second stage of PRISMA, or the *screening* stage, looks at excluding inappropriate and unrelated papers. This stage reduced the corpus count to 202. Search applied the following inclusion criteria:

- *Language:* Only articles written in English language were included. This step omitted 67 articles.
- *Keywording:* Only articles with subject-relevant keywords coded by Scopus for indexing purposes (also known as *Indexed Keywords* by Scopus) were included. Subject-relevant keywords included, and are not limited to: *Education, University, Higher Education, Learning Environment(s), Learning System(s), E-learning, Online Learning, Education Computing, Intelligent Tutoring System(s), Computer Aided Instruction, Learning Analytic(s), Curricul(a/um), Teaching, Learning, Learning Process(es), Collaborative Learning, Student(s), Academic Performance(s), Ethic(s), Ethical Consideration(s), Ethical Issue(s), Ethical Technolog(y/ies), Fairness, Data Privacy, Trust, Moral(s), Moralit(y/ies), Perception(s)*. Excluded keywords included, and are not limited to: *Medical Ethic(s), Bioethic(s), Clinical Stud(y/ies), Patient Simulation(s), Doctor-Patient Relationship(s), Risk Assessment(s), Human Resource*

Management, Software Engineering, Electronic Assessment(s). This step omitted 7,805 articles.

- *Publication Stage:* Only peer-reviewed final articles published in scientific venues (e.g., books, journals and conferences) were included, for rigority of selection. This step omitted 62 articles.
- *Year of Publication:* Only articles published in 2018 and beyond were included, to ensure recency of literature. Rigorous peer-reviewed articles would have reviewed key prior related literature within their respective papers. This step omitted 67 articles.

The third stage of PRISMA, or the *eligibility* stage, requires scanning title and abstracts, and full papers to identify relevant eligible articles. This stage yielded a final corpus count of 33 articles. Search applied the following inclusion criteria:

- *Assess Titles and Abstracts for Suitability:* Only relevant titles and abstracts were included. There should be explicit and direct references to the subject matter. This step omitted 132 articles.
- *Assess Full Papers for Suitability:* Only relevant full papers were included. An additional inclusion criterion here was that all articles should have their full text accessible for analysis. This step omitted 37 articles.

A summary of the PRISMA approach is shown in Fig. 2.

Data Extraction

As a citation engine, data in Scopus is highly structured and robustly tagged, delivering metadata for analytical purposes, including (i) author(s), (ii) document title, (iii) affiliation(s), (iv) year, (v) publication, (vi) volume, issue and page source, (vii) citation, (viii) document type, (ix) keywords, and (x) digital object identifier (DOI), among others.

The final pool of 33 primary studies were thoroughly analyzed to answer the research questions of this study (refer to Appendix). Information that was extracted from Scopus included: (i) citation information, such as author(s), title, year, publication, and citation count etc., (ii) bibliographical information, such as affiliation(s), and publisher etc., (iii) abstract, (iv) keywords, and (v) references.

Classification and Analysis

Using the data extracted from Scopus, the study utilized *Tableau Desktop Professional* version 2021.1.20 to perform exploratory data analyses to address RQ1. Tableau platform allows powerful conversion of complex computations into appealing data visualizations.

With the Scopus extracted data, research utilized a corpus analysis platform *CorTexT* (Breucker et al., 2016) to perform text parsing, and a first pass of topic modelling and network mapping, so as to identify major thematic representations of corpuses comprising of Author Keywords and Indexed Keywords. This allowed us to perform machine learning for pattern recognition, utilizing unsupervised text mining techniques on these keywords to identify useful patterns.

Using the Python Library *pyLDAvis* (Sievert and Shirley, 2014), topic modelling generated a topic representation of the keyword corpus' textual fields using the Latent Dirichlet Allocation method, which allowed a visualization of the most relevant words fitting to the topic. Here, each topic was defined as a keyword probability distribution, and each document was defined as a topic probability distribution. Given the total number of topics defined, the topic model was inferred by probabilistically assigning topics to documents, and positioned in 2D according to a multi-dimensional scaling algorithm for visualization purposes.

While topic modelling provided a sense of the latent themes from the underlying keywords, research further performed network analyses to visualize thematic keyword representations in a clustering format, where each keyword was grouped with distinct members, and linked via proximity measures. The Louvain hierarchical community detection algorithm was used (Aynaud, 2020). This algorithm is based on modularity optimization, where the optimal linkage densities are measured, taking into account within-cluster and between-cluster linkages. Louvain algorithm is efficient on large networks.

The first pass of topic modelling and network analyses above allowed the identification of distinct sub-themes of AI application areas and ethical issues. With the key sub-themes of AI application areas and ethical issues identified as a priori, each article was thoroughly evaluated and coded to classify the following: (i) application areas where AI is used in assessments (e.g., assessment curation and personalized feedback etc.), and the (ii) type of ethical issues relevant to AI-based assessments as cited in paper (e.g., fairness and explainability etc.). This would allow us to address RQ2 and RQ3.

The study then undertook a thorough review of the full papers, and provided further analyses to tabulate the following: (i) breakdown of each type of ethical issues identified in each paper (e.g., how explainability of AI systems is an important ethical consideration in assessments etc.), and (ii) breakdown of mitigation and intervention methods for each ethical issue as highlighted in each paper (e.g., applying data sanitization to reduce risk of discriminatory decision making from AI systems etc.).

Using the coded sub-themes of AI application areas and ethical issues, research undertook the second pass of topic modelling and network analyses. The topic modelling and network analyses outputs would be used to guide the identification of the major research themes to address RQ4.

Evaluation of Validity

In the application of systematic literature mapping, it was useful to consider the following types of validity to ensure that the methodology was robustly constructed. These included (i) descriptive validity, (ii) interpretive validity, (iii) theoretical validity, and (iv) generalizability (Petersen and Gencel, 2013). Detailed reporting of the systematic mapping methodology process, including the evaluation of validity, helps improve repeatability of the study.

- *Descriptive validity*

  This describes the extent to which there existed objective and accurate observations. To lower the risk of this threat, a data extraction and coding spreadsheet was designed to support data recording. This provided objectification of the data extraction process, and

allowed interventive correction to ensure accuracy, if required. As such, this risk was considered under control.

- *Interpretive validity*

  This describes, given the data extracted and coded, the validity of the conclusions drawn. A key threat might be researcher biasness. This was alleviated by ensuring that no primary papers authored by the authors were included in the primary papers extracted, which reduced threats in interpretation.

- *Theoretical validity*

  This describes the prospect of being able to capture what was purported to be captured. Research looked to ensure that the thematic phenomena identified in the paper represented the patterns of the real world. Scopus provided a strong integration with major publishers, and its wide interdisciplinary focus ensured the lowering of probability of missing key research information. In terms of paper screening, careful curation of keywords, selection of final peer-reviewed papers published in scientific venues, and the recency of literature ensured that the literature reviewed were accurate, peer-reviewed and timely. Extensive in-depth reviews were also made in the full texts to ensure that each paper included were suitable. For quality assessment and to reduce potential biasness, the methodology and data extraction process were checked by an independent external reviewer, with subject matter-relevant background.

- *Generalizability*

  This describes the external validity (i.e., generalizability on the basis of repeatability and extendibility of results from this paper to other research), and internal validity (i.e., causal effect between application of AI in assessments and their related ethical issues). In the presence of a wide range of similar ethical discourse on different AI applications, the classification and analysis methodology should not result in major threats for both internal and external validity. However, it is acknowledged that external validity may be influenced by factors such as domain-specificity (Leslie, 2019), cultural differences (Awad et al, 2018), and sample size (Khan et al., 2022), and future studies can help alleviate external validity concerns.

Limitations

Although Scopus is a robust database of digital records for peer-reviewed literature to map and survey specialized scientific areas (e.g., Fahimnia, Sarkis and Davarzani (2015), Rodrigues et al., (2014)), we recognize that Scopus is not the only one available. There are other valid alternatives, including Web of Science, ACM, IEEE Xplore, EBSCO Host, Wiley, SAGE Journals, and Taylor and Francis, among others. Among these databases, there are arguments brought forth by Campedelli (2021) that the degree of overlap between the publication titles in both databases may be closer to 50% to 60%, hence including both databases may have value. However, the discriminant feature that supported the selection of Scopus is that its informative tagging of all papers by professional indexers using Indexed Keywords existed at a higher frequency and provided a richer pool of content for each item, in particular, for textual mining purposes, as compared to Keyword Plus from Web of Science. As a first assessment on this subject matter, this should suffice. For follow up works, it will be useful to consider integrating

Web of Science and/or other relevant databases to provide a comprehensive scan of this landscape.

Secondly, throughout this chapter, there were no assumptions made on the intrinsic value of thematic diversity. This work utilizes unsupervised machine learning techniques to search for latent topics embedded in existing literature, and through which seeks to act as a key building block upon which future research can be applied. The study of possible inhibited or dysfunctional states within this thematic diversity, plausibly due to scholarly or technology inertia, or the lack of infrastructure or skills resulting in resistance to state-of-the-art adoption, are outside of the scope of this study. It may be useful for future research to quantify the value of this thematic diversity, in terms of (i) its operationalization impact as a segregable assessment pipeline component, and (ii) the extent to which ethics (or the lack thereof), either in isolation or in combination, impact upon the assessment pipeline component. Drawing from Klinger, Mateos-Garcia and Stathoulopoulos (2020), it may be useful to apply Weitzman (1993)'s economic valuation of ecological diversity, taking into account the cost-benefit analysis of preserving diversity and the threshold below which the archetypical research theme becomes unsustainable.

FINDINGS

This section presents the findings based on an analytical investigation of selected published primary papers identified as relevant to the study.

Research Question (RQ) 1: Where do the studies that discuss ethical issues relating to AI-based assessments arise from?

To address RQ1, research undertook exploratory data analyses to explore where the studies discussing ethical issues relating to AI-based assessments arise from.

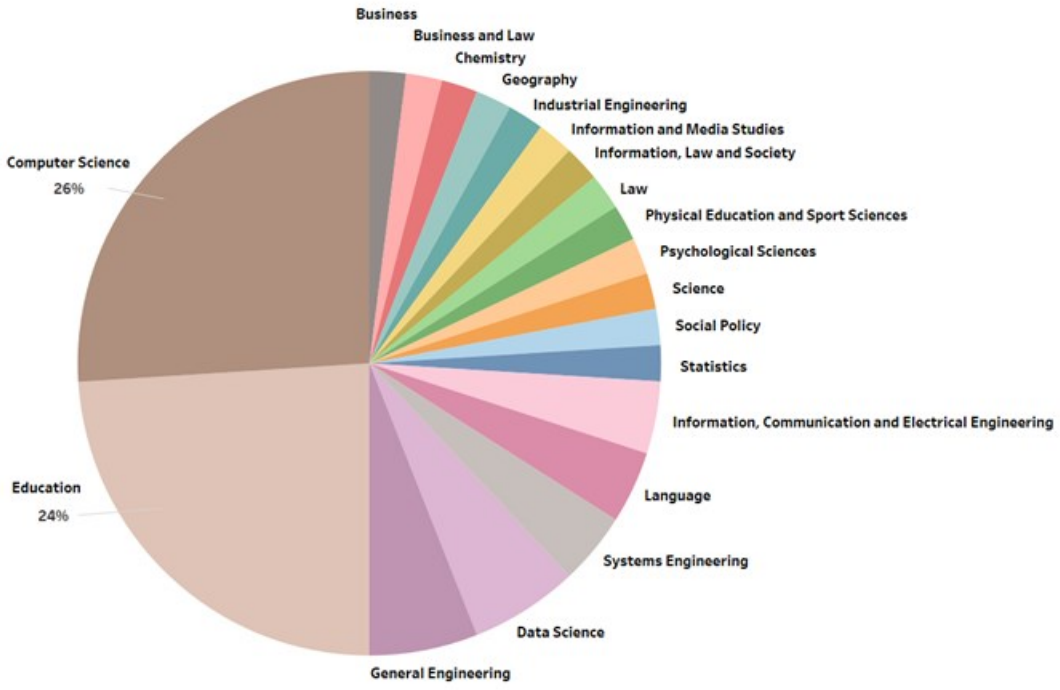

Figure 3. Breakdown of disciplines tied to authors' affiliated department.

In Fig. 3, it was noted that, through observing the authors' affiliated departments, the papers emerged from researchers across a wide range of disciplines. About half of the researchers originate from Computer Science and Education departments, with Computer Science leading marginally. The remainder stemmed from eighteen disciplines, ranging from soft pure disciplines such as Language, to hard applied disciplines like Engineering. This suggests that research pertaining to AI-integrated assessments and ethics do not necessary emanate from the Education departments, but from a broad spectrum of disciplines.

In Fig. 4, it was noted that most papers originate from The University of Melbourne, Australia (4); Carnegie Mellon, United States (3); University of Oulu, Finland (3); and University of Eastern Finland, Finland (3). The remainder are institutions that published two papers; the yellow circles denote institutions that published one paper. Papers from The University of Melbourne originated from Law, Psychological Sciences, and Engineering schools. Papers from Carnegie Mellon University originated from the Language Technologies Institute, Eberly Center for Teaching Excellence, and Human-Computer Interaction Institute. Papers from Oulu University originated from the Geography, Chemistry and Education departments. Papers from University of Eastern Finland originated from the Computing and Education Sciences departments. These provided an overview of the disparateness of affiliation and departments.

Subject matter is a complex multi-faceted issue that spans, among others, pedagogy, technology, and psychology domains. Taking cues from present research work, it may be useful for the nature of such research to have more pluralistic cross-disciplinary collaborations (e.g., computer science, education, social science etc.) in research and development work, to achieve sounder theoretically underpinned methodology approaches and more stakeholder inclusivity (Raji, Scheuerman and Amironesei, 2021).

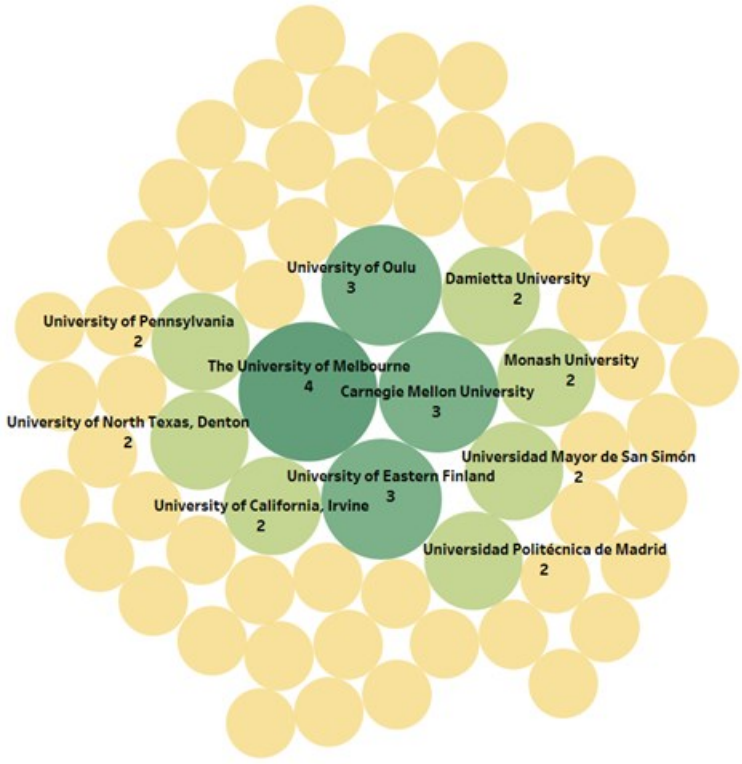

Figure 4. Breakdown of affiliated institutions

In Fig. 5, it was noted that the top five countries where the papers originated were the United States, Australia, Brazil, Finland, and Spain (tied at fourth). Bozkurt (2020) found that countries that have historically dominated in educational technology research stemmed from the United States, United Kingdom, and Taiwan. In a different sub-domain in education, ethics-related education, for instance, in the field of healthcare, is also dominated by United States and Taiwan (Andersson et al., 2022). From these lists, aside from the United States which continued to lead in publication numbers, United Kingdom and Taiwan appear to be laggards. There are still room for different countries to play leading roles in this research aspect, as innovators, early adopters and early majority.

In Fig. 6, it was noted that the top publication venues were Computer Science venues, namely (i) 4 conference papers published in Springer's *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, (ii) 3 conference papers published in Association for Computing Machinery's (ACM) *ACM International Conference Proceeding Series*, and (iii) 3 journal papers published in Institute of Electrical and Electronics Engineers' (IEEE) *IEEE Access*.

As a proportion, 55% of the papers were published in Computer Science venues such as those listed above, 24% were published in Educational Technology venues, such as the *British Journal of Educational Technology, International Journal of Artificial Intelligence in Education*, and *Journal of Learning Analytics*; and 15% were published in Education journals (mainly domain-specific education) such as the *International Journal of Information and Communication Technology Education, Journal of Information Systems Education*, and *International Journal of Engineering Education*. This suggests that Computer Science venues play a leading role (and may hold suitable target audience who can and are interested to participate) in the discourse of ethics in AI-based assessment practices.

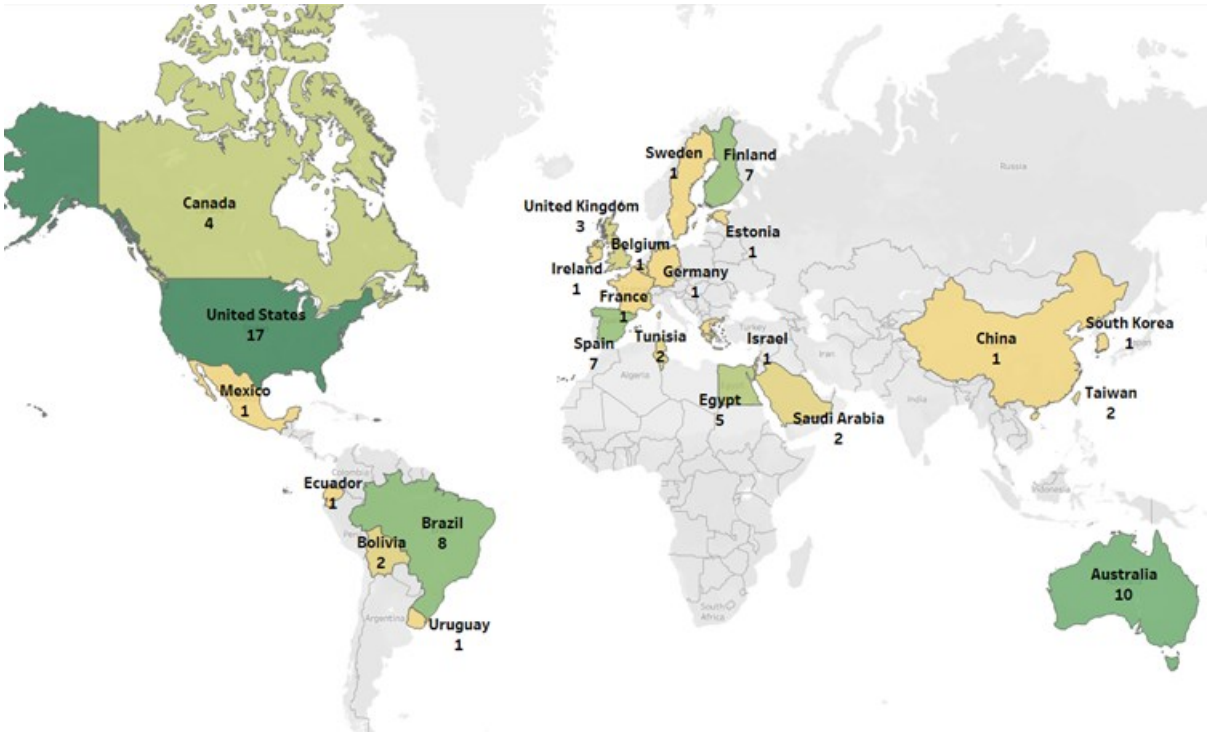

Figure 5. Breakdown of author locations

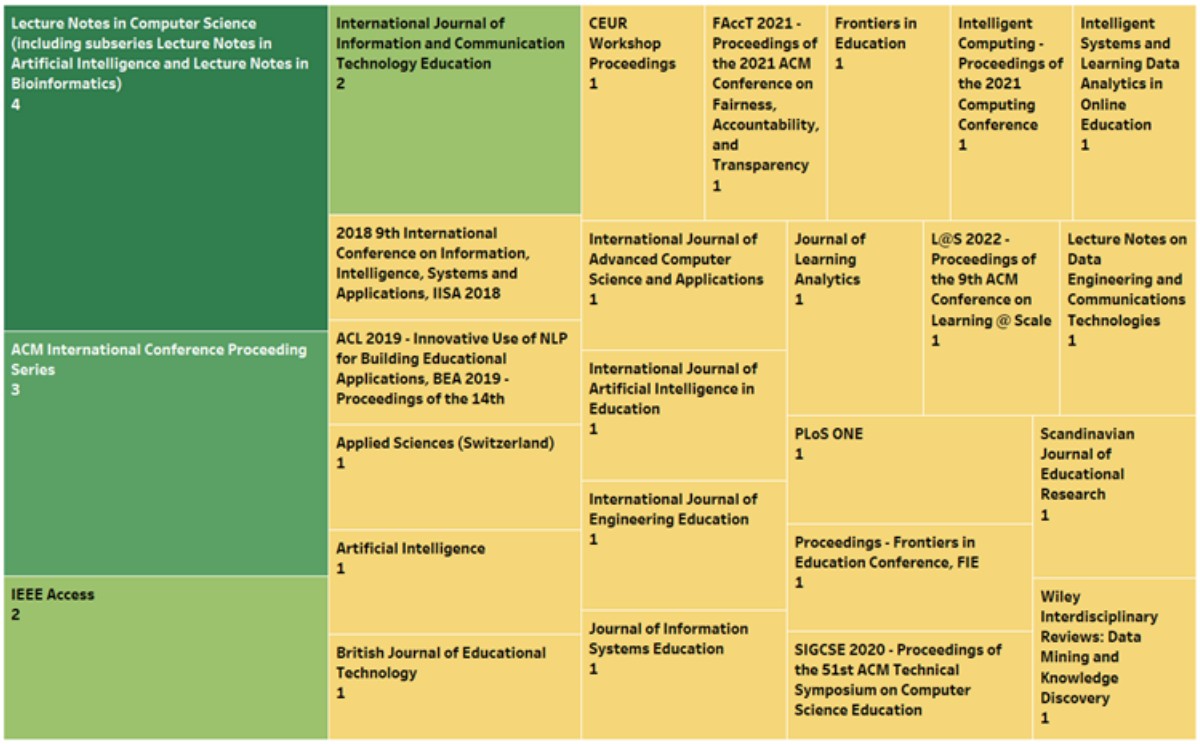

Figure 6. Breakdown of publication source

In summary, the landscape is still at the early phases of the technology adoption curve, with enough room for researchers beyond the education domain. Present research is marginally led by the computer science discipline. However, research in this area is highly interdisciplinary, spanning from soft pure to hard applied disciplines, including but not limited to, education, computer science, educational and technology philosophy and psychology, social policy, and law. When applied on specific domains, it may also involve domain specific knowledge, such as sport science. Hence, a cross-pollination of ideas through inter-departmental collaborations can be highly beneficial to advance in this field.

First pass of topic modelling and network analyses

Topic modelling was performed, where the optimal number of topics were generated using a model with the highest topic coherence. Further, we performed network analyses to identify topic clusters. These allowed us to recognize patterns in an unsupervised machine learning approach.

From this first pass of topic modelling, ten latent topics were identified. For instance, in Fig. 7, we observed the latent topic of AI modelling and predictive analytics that might be relatively more closely linked to ethical issues of explainability and fairness. This mirrored well with the network analyses visualization in Fig. 8. Here, we observed a more granular fourteen latent topic clusters, with the clusters of *Modelling*, and *Predictive Analytics and At-Risk Student*s, situated in close proximity in the cluster diagram. The higher granularity of the outputs allowed us to identify distinct sub-themes of AI application areas and ethical issues. The top keywords and latent topics of topic modeling are shown in Table 1.

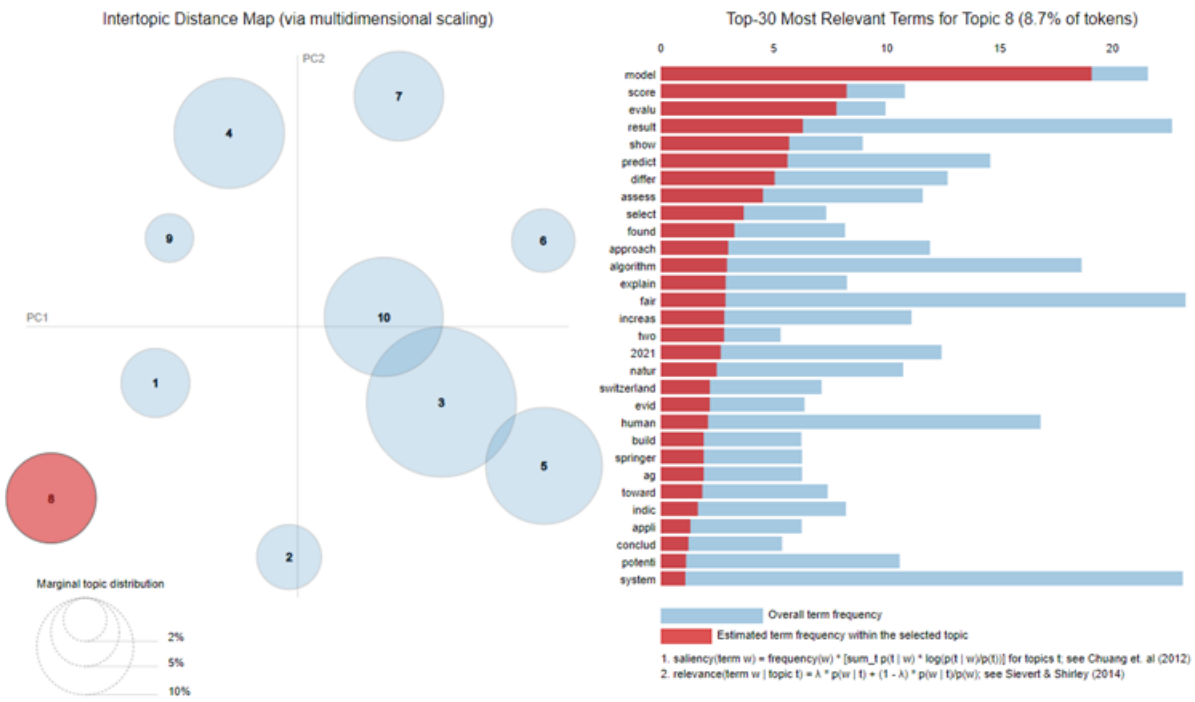

Figure 7. Topic modelling of keyword corpuses

Latent Topic	Percentage of Tokens	Top Keywords
Engineering systems	5.0%	Build; Design; Model; Technology; System
Automated grading	14.5%	Feedback; Response; Evaluation; Grading; Automation
Intelligent tutoring and feedback	23.9%	AI; Teacher; Intelligent; Support; Feedback
Predictive analytics and at-risk students	13.0%	Predict; Improve; Support; Measure; Environment
Explainable AI	4.5%	Explainable; Fair; Ethic; Algorithm; Review
Forecasting	4.3%	Forecast; Model; Automation; Learning; Predict
Adaptive robots	4.3%	Adapt; Environment; Tool; Robot; System
Modeling	8.7%	Model; Score; Evaluation; Result; Predict
Security and cheating	2.50%	Privacy; Data; Institution; Risk; Ethics
Assessment tasks	15.10%	Curation; Delivery; Task; Construct; Generate

Table 1. First pass of topic modelling – Latent topic and top keywords

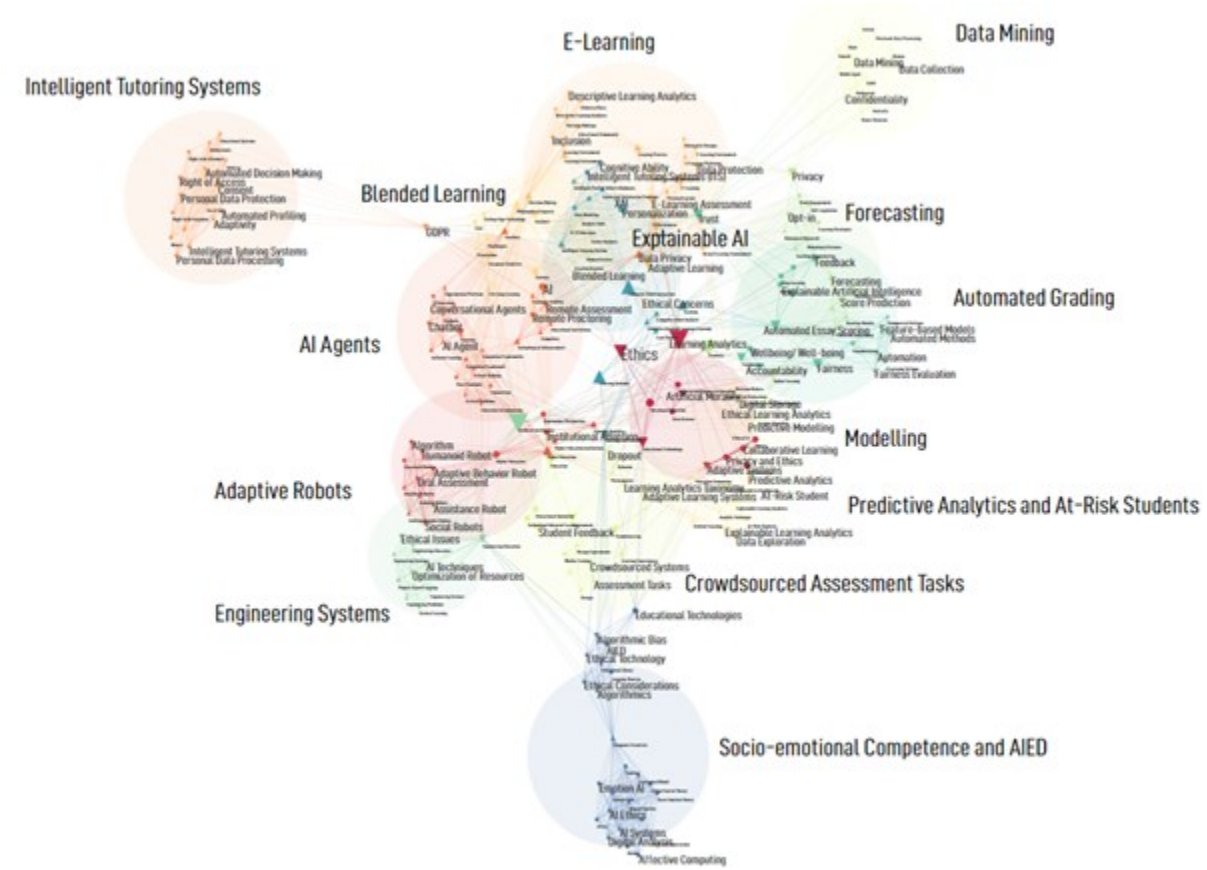

Figure 8. Network analysis of keyword corpuses

Through the review of the first pass of topic modelling, network analyses outputs, and full paper reviews, the study extensively identified fourteen sub-themes of AI application areas and ten sub-themes of ethical issues. We populate them in Tables 2 and 3, respectively.

Research Question (RQ)2: What are the main AI use cases relating to assessments?

As shown in Table 2, the fourteen sub-themes of AI application areas were, namely:

- *AI-based assessment construction, curation, or delivery*

  AI can be a useful tool to:

  - construct assessments organically through, e.g., the use of generative AI to generate draft question samples for an assessment practitioner's review (Last and Danon, 2020);
  - construct assessments collaboratively through e.g., crowdsourcing of assessment tasks (Ahn et al., 2021); and
  - curate and deliver personalized formative assessment (Gupta and Chen, 2022; Heo & Lee, 2019) through e.g., curating multiple choice format formative assessments generated via AI conversational agents (Pereira, 2016).

  This form of AI utilization is cited in 45% of the primary studies. This is the third highest cited AI utilization form.

- *AI-based socio-emotional assessment*

  AI can be used to:

  - assess non-cognitive psycho-emotional behavior qualities, such as persistence and grit, initiative and adaptability etc., through ambient intelligence (Stark and Hoey, 2021; Williamson, 2021); and
  - discover socio-emotional patterns that may be predictive of assessment performance (Peña-Ayala, 2018).

  This form of AI utilization is cited in 6% of the primary studies.

- *AI-based group assessment using collaborative analytics*

  AI can be used to:

  - assess collaborative dynamics in group projects through the application of analytics on learners' rich multimodal interaction data; and
  - model interdependent relationships to support students' collaboration dynamics through adaptive methods (Schneider, Dowell and Thompson, 2021).

  This form of AI utilization is cited in 3% of the primary studies.

- *AI-derived opportunities for learning intervention or assistance*

  AI can be useful for:

  - identifying and supporting learning intervention opportunities in formative assessments (Shabaninejad et al., 2021; White et al., 2021); and
  - providing scaffolding assistance in assessments, e.g., with the help of AI-driven hints through Intelligent Tutoring Systems (Conati et al., 2021; Latham and Goltz, 2019).

  This form of AI utilization is cited in 52% of the primary studies. This is the second highest cited AI utilization form.

- *AI-generated personalized feedback*

  AI can be a useful tool to personalize feedbacks in terms of:

  - clarifying approaches to attempt a new assessment;
  - reviewing assessment performance; and/or
  - recommending takeaways from assessments to improve learners' demonstration of competence at their current or future workplace (Gupta and Chen, 2022; Merikko et al., 2022).

  This form of AI utilization is cited in 55% of the primary studies. This is the highest cited AI utilization form.

- *AI-based predictive analytics*

AI can be a useful tool to predict assessment outcomes, to support educators' focus on learning scaffolding (Chounta et al., 2022; Kim et al., 2021).

This form of AI utilization is cited in 15% of the primary studies. This is the fifth highest cited AI utilization form.

- *AI-based teaching evaluation*

  The term '*assessment*' in an educational context is extended to the assessing of educators as a formal evaluation process to review teaching effectiveness and perform in classrooms (Tlili et al., 2018).

  This form of AI utilization is cited in 9% of the primary studies.

- *AI-facilitated response and grading*

  AI can be a useful tool for:

  - adaptive understanding and responding to learners in formative assessments (Khairy et al., 2022); and
  - automated grading of structured (e.g., multiple choice questions) and non-structured (e.g., open-ended or essay questions) assessment responses (Litman et al., 2021; Kumar and Boulanger, 2020).

  This form of AI utilization is cited in 30% of the primary studies. This is the fourth highest cited AI utilization form.

- *AI-based proctoring*

  AI can be a useful tool to proctor on-site (e.g., in classrooms or exam halls) or remotely (e.g., online assessment on Massive Open Online Courses, or MOOCs) to lower the risk of cheating in quizzes and exams (Elshafey et al., 2021).

  This form of AI utilization is cited in 6% of the primary studies.

- *AI-based authentication and security gateways for the conducting of assessments*

  AI can be used to build authentication and security measures in assessment systems, so that the risk of personal and educational data leakage and exploitation can be lowered (Kiennert et al, 2019).

  This form of AI utilization is cited in 3% of the primary studies.

- *AI-backed plagiarism detection in assignment submissions*

  AI can be a useful tool to perform plagiarism detections, so that assignment submissions that present someone else's works or ideas without acknowledgement or consent can be detected (Kiennert et al, 2019).

  This form of AI utilization is cited in 3% of the primary studies.

- *AI-integrated communication dashboard for assessment outputs*

An AI-integrated communication of assessment outputs can be designed in an inclusive format that promotes:

- clear automated communication and feedback of assessments, accounting for the nuanced sensitivity of communications that may be absent in poorly designed AI-based text generators or audio-visual responses; and
- automation of inclusiveness actions. For instance, for learners with disabilities (such as motor disabilities), there may be an automation of increased time limit catered to complete an assessment (Costas-Jauregui et al., 2021).

This form of AI utilization is cited in 3% of the primary studies.

- *Autonomous intelligent agents for assessment purposes*

  This relates to the care given to the design of AI algorithms to create intelligent agents that make autonomous decisions for assessment purposes (Casas-Roma and Conesa, 2021; Hakami and Hernández-Leo, 2020).

  This form of AI utilization is cited in 9% of the primary studies.

- *AI-based behavioral tracking for assessment purposes*

  AI can be used to track spatio-temporal data of assessment behavior:

  - to predict learning intervention opportunities; and/or
  - for assessment evaluation purposes (White et al., 2021).

  This form of AI utilization is cited in 3% of the primary studies.

Paper	Applications of AI Related to Assessments														Note
	Construction, curation, or delivery	Socio-emotional assessment	Group assessment	Learning intervention/ assistance	Behavioral tracking	Personalized feedback	Predictive analytics	Teaching evaluation	Grading and response	Proctoring	Authentication and security	Plagiarism detection	Communication dashboard	Autonomous intelligent agent	
Gupta and Chen (2022)	•														- Telegram chatbot @dawebot, an AI-enabled conversational agent, provides formative assessments via multiple choice questions in any subject (Pereira, 2016)
Chounta et al. (2022)	•					•	•								-
Deho et al. (2022)				•			•								-

Shabaninejad et al. (2022)				•		•									-
Nazaretsky, Cukurova and Alexandron (2022)	•			•		•									-
Pontual Falcão et al. (2022)				•		•									-
Merikko et al. (2022)				•		•									-
Khairy et al. (2022)	•														-
Megahed, Abdel-Kader and Soliman (2022)									•						-
Conati et al. (2021)	•														- AI-driven hints on assessments delivered through Intelligent Tutoring System
González-Calatayud, Prendes-Espinosa and Roig-Vila (2021)	•			•		•			•						-
White et al. (2021)					•										- Tracking data of assessment behaviour applied for learning intervention
Ahn et al. (2021)									•						-
Stark and Hoey (2021)		•													-
Papa and Jackson (2021)														•	-
Kim et al. (2021)						•									-
Litman et al. (2021)									•						-
Casas-Roma and Conesa (2021)				•		•		•	•					•	-
Costas-Jauregui et al. (2021)				•		•							•		-
Elshafey et al. (2021)										•					-
Schneider, Dowell and Thompson (2021)			•												- Application of collaborative analytics in group assessments
Gedrimiene et al. (2020)	•			•		•									-

Kumar and Boulanger (2020)						•	•		•						-
Khosravi, Sadiq and Gasevic (2020)	•														-
Martín Núñez and Lantada (2020)	•			•		•			•						-
Hakami and Hernández-Leo (2020)				•		•		•	•					•	-
Mougiakou, Papadimitriou and Virvou (2019)	•			•		•			•						-
Mayfield et al. (2019)	•			•		•			•						-
Latham and Goltz (2019)	•			•		•									- AI conversational agent - AI-driven hints on assessments delivered through Intelligent Tutoring System
Tlili et al. (2019)	•			•		•									-
Kiennert et al. (2019)										•	•	•			-
Peña-Ayala (2018)	•	•		•		•	•								-
Tlili et al. (2018)	•			•		•		•							-
Total Count	**15**	**2**	**1**	**17**	**1**	**18**	**5**	**3**	**10**	**2**	**1**	**1**	**1**	**3**	
Percentage	*45%*	*6%*	*3%*	*52%*	*3%*	*55%*	*15%*	*9%*	*30%*	*6%*	*3%*	*3%*	*3%*	*9%*	

Table 2. Breakdown of sub-themes of AI application areas by paper

Research Question (RQ)3: What are the main ethical issues arising from the AI implementations relating to assessments?

As shown in Table 3, the ten sub-themes of ethical issues were, namely:

- *Inclusivity*

  This ethics principle relates to inclusive and accessibility considerations applied to AI systems to meet different student needs in a personalized environment at scale.

  - Inclusiveness is concerned about exhibiting empathy towards sensitive learner conditions, such as health, disabilities and learning disorders (e.g., pregnancy, visual handicap, or dyslexia), gender, race, prior education backgrounds (e.g., non-native speakers), and socio-economic backgrounds.

In the design of the AI system, actions are taken to respect the diversity of learners and ensure that prejudices, stereotypes, discrimination and biasness do not creep into assessments (Gupta and Chen, 2022; Martín Núñez and Lantada, 2020; Tlili et al., 2019).

- Inclusiveness is concerned about considering the sensitivity of communication and feedbacks generated by AI systems, so that learners are not negatively impacted by AI-generated textual comments or audio-visual responses when they clarify approaches to a new assessment or receive takeaways from completed assessments (Costas-Jauregui et al., 2021).
- Inclusiveness is concerned about lowering the chances of conformity, peer pressure and segregation that may be reinforced because of AI generated decisions, which can negatively impact both educators and learners (Gedrimiene et al., 2022).

This ethics principle is cited in 18% of the primary studies.

- *Fairness*

This ethics principle relates to fair, equitable and appropriate assessment practices that should be perpetuated by AI systems. It is noted that the definition of fairness is plagued with problems of subjectivity, contextualization and cultural-specificity (Hakami and Hernández-Leo, 2020).

- Fairness is concerned about the treatment of data and algorithmic bias to ensure diversity, equity, non-prejudice and non-favoritism towards learners' sensitive attributes, so that needs of minority groups are not disadvantaged or underrepresented. This overlaps with the concept of inclusivity. (Gupta and Chen, 2022; Deho et al., 2022; Megahed, Abdel-Kader and Soliman, 2022; Casas-Roma and Conesa, 2021; Latham and Goltz, 2019; Tlili et al., 2019).

  Tied to this discourse are the ethical concepts of (Mayfield et al., 2019):

  (i) *Allocation harm:* This relates to the equitable distribution of resources of learning, such that the possibility of differential outcome distributions generated by AI systems are minimized.

  (ii) *Representational harm:* This relates to the stereotyping biasness perpetuated by data and/or algorithm, resulting in the marginalizing of groups of learners.

- Fairness is concerned about the unintended labelling or profiling of learners, which can affect their learning journey and well-being (Peña-Ayala, 2018).
- Fairness is concerned about, in the context of socio-emotional assessments, how applying universal assumptions on emotional states are harmful, due to different cultural context of emotional interactions and norms (Stark and Hoey, 2021).
- Fairness is concerned about the ad hoc implementation of AI systems, in the absence of:

(i) standard code of practices and ethics; and

(ii) befitting monitoring and accounting mechanisms,

which may impact the implementation of fair, equitable and appropriate assessment practices (Tlili et al., 2018).

This ethics principle is cited in 42% of the primary studies. This is the joint-second highest cited ethical issue, alongside explainability.

- *Accountability*

  This ethics principle relates to the responsible discharge of AI ethics when designing and delivering AI systems, depending on the roles and contexts, in a consistent manner. Stakeholders who construct, operate and use AI systems should be accountable for AI systems and decisions.

  - Accountability is concerned about the moral obligation for institutions to reflect and act, given that it has access to data that may know and understand how students learn (Costas-Jauregui et al., 2021). Students are data subjects who are not generally able to influence the handling of data in an ethical manner (Gedrimiene et al., 2020).

    Care should be applied when overseeing sensitive data. It was noted that between 2007 and 2011, there were 133 incidents linked to educational institutions unintendedly disclosing sensitive learner information (Stiles, 2012). Such incidents can lead to reputational, legal and/or financial liabilities (Tlili et al., 2018).

  - Accountability is concerned about the processes where relevant stakeholders provide reasons and take responsibilities for the actions of decisions influenced by AI algorithms (Hakami and Hernández-Leo, 2020). There should be proper consent, and non-maleficence academic interventions.

    For instance, in the development of AI systems for socio-emotional assessment, designers of AI systems should recognize and be accountable to the fact that there may exist diverse human attitudes to emotions to ensure fairness. Further, AI developers should also be cognizant that care should be applied to the underlying data, as data can be misused and emotions can be harnessed as a social phenomenon (Stark and Hoey, 2021).

  - Accountability is concerned about demonstrating compliance with relevant regulations and guidelines (Latham and Goltz, 2019). However, it is noted that there may exist challenges relating to such compliance.

    For instance, there may exist a lack of interoperability of regulatory guidelines on misuse of private information (e.g., European Union and Latin America), and a lack of clarity regarding whether if the institution or the students own the data that are shared by the students (Costas-Jauregui et al., 2021).

    Furthermore, for online courses offered worldwide, informed consent for use of data can be affected by the data protection regulations where the learner is

domiciled. This greatly increases the difficulty of compliance efforts (Tlili et al., 2018).

- Accountability is concerned about the availability of avenues for redress that are fair and unprejudiced due to the adverse use of AI systems to account for detrimental individual or societal effects.

This ethics principle is cited in 30% of the primary studies. This is the fourth highest cited AI utilization form.

- *Accuracy*

  This ethics principle relates to the reliability and validity of assessments when an AI system is applied. In the presence of possible biasness or errors introduced by data and AI algorithms, which may compromise the reliability and validity of assessments, there should exist measures to establish, log, communicate, diagnose and mitigate the biasness or errors.

  - Accuracy is concerned about poor data quality.

    For instance, due to the distributed nature of online learning, data collected may be incomplete or erroneous. Linked to this latter notion is the right for data subjects to check and rectify data collected, so that data inputs are accurate.

    Poor data quality can negatively impact AI-driven decisions (Tlili et al., 2018).

  - Accuracy is concerned about inappropriate data inputs.

    For instance, predictive models with imbalanced dataset (e.g., gender) may generate less effective predictions (e.g., for minority gender). The possibility of discriminatory and unfair practice extends to other socio-economic demographical information, such as ethnicity, underrepresented groups etc. (Chounta et al., 2022).

    In another example, in the implementation of collaborative analytics, it is vital to capture multimodal data that best represents and assesses the collaborative interactions of students, otherwise the AI-generated outputs cannot be relied upon (Schneider, Dowell and Thompson, 2021).

  - Accuracy is concerned about inaccurate understanding and interpretation of learner responses in assessments.

    For instance, adaptable humanoid robot may not understand or interpret the responses of learners in an oral assessment well. Correct answer rate, if affected by learner's pronunciation or robot's lack of contextual understanding, can affect confidence in the AI system (Khairy et al., 2022).

    Research by Ahn et al. (2021) also showed that automated grading of learners' work, which contains complex data, rich semantic meaning and idiosyncratic and local nuances, may not be well graded by present computational approaches that utilize metrics such as counts of parts of speech and essay length as proxies for writing complexity and quality.

Further, present systems are generally rigid in formulation of tasks and grading. For instance, the rejection of lexicon and grammar of minority dialects. This limits the choice of tasks, types of acceptable answers and styles of writing (Mayfield et al., 2019).

- Accuracy is concerned about the validity and reliability of an assessment instrument.

  Kim et al. (2021) shared that, to allow for shorter assessments in online learning, it is imperative that the question set reduction guided by AI algorithms is done in such a way where the reduced set is able to approximate the original assessment's evaluation of learning. This can help ensure reliability and trust on assessment instrument.

- Accuracy is concerned with the treatment of prediction errors and biases.

  For instance, in the case of cheating detection, the treatment of non-cheating cases that were falsely detected as cheating (or false positives), and cheating cases that were not detected (or false negatives). The false positives and negatives will have to be reduced to improve accuracy rates (Kiennert et al., 2019) .

- Accuracy is concerned about the possibility of inaccurate predictions, due to the "gaming" of AI systems.

  For instance, there exist a possibility for students to modify their behavior and "game" the AI system, when they have the knowledge that they are assessed by the AI system, and the knowledge of the parameters of the AI model. This can create inaccurate AI decisions when assessing students (Tlili et al., 2018).

This ethics principle is cited in 27% of the primary studies. This is the fifth highest cited AI utilization form.

- *Auditability*

  This ethics principle relates to the permitting of independent third-party reviewers to audit, analyze and report findings relating to the usage and design of data and AI algorithms in assessments.

  - Auditability is concerned about the understanding, validating, reviewing and improving of the AI system applied, so that there are appropriate transparency, traceability and utilization of data and AI algorithms, and appropriate validity and reliability of assessment instruments. However, it is noted that challenges may arise if algorithms are proprietary (Tlili et al., 2019; Casas-Roma and Conesa, 2021).

  This ethics principle is cited in 9% of the primary studies.

- *Explainability*

  This ethics principle relates to the lowering of opacity relating to data, AI algorithms and AI-driven decisions, the justification of its use, and the communication of details

in a non-technical easy-to-understand manner to relevant stakeholders (Kumar and Boulanger, 2020; Casas-Roma and Conesa, 2021).

- Explainability is concerned with transparency of the design of AI systems. Transparency is tied to information availability, accessibility conditions, possibility of pragmatic decision-making assistance, and user knowledge (Nazaretsky, Cukurova and Alexandron, 2022).

  It is importance for AI-based assessments to be developed in an explainable and transparent manner to safeguard trust and fairness with human stakeholders. For instance, for AI recommender systems, why are some assessment questions recommended over others (Chounta et al., 2022). "Black-box" AI recommendations, which provide low or no insights into recommendation rationales (Abdi, 2020), may be plagued with biases and confounding problems (Bastani, Bastani and Kim, 2018; Khosravi et al., 2021), resulting in unjustified actions and discrimination (Papa and Jackson, 2021).

  However, it is noted that full transparency may be harmful, as users can "game" the system to their benefit and the detriment of others (Hakami and Hernández-Leo, 2020). Furthermore, challenges may arise from disclosure of proprietary algorithms or trade secrets (Latham and Goltz, 2019).

- Explainability is concerned about the use of explanations to gain insights into the behavior of AI systems.

  Present explainability approaches include global and local approaches. The former synthesizes and uncovers qualities of inputs that affects model behaviors on a global basis, whereas the latter looks to explain the model's behavior to a specific input. Another more recent approach is to leverage on generative capabilities of models to self-explain a human-understandable explanation for input-output responses (Bommasani et al., 2021).

- Explainability is concerned about tradeoffs relating interpretability and complexity of AI systems.

  The design of the AI algorithms may be potentially complex. For instance, for collaborative analytics, the use of interdependent modelling to assess group dynamics and outcomes, may raise challenges on explainability of AI systems. Interdependent models (which looks at students' influences on one another over time) may be significantly more complex than independent models (which looks at students as isolated events). There is a need to assess the trade-off between model complexity and explainability, and ascertain if simpler models are sufficient to model dynamic interdependence (Schneider, Dowell and Thompson, 2021).

  Deho et al. (2022) suggests that interpretable models (e.g., logistic regression) may provide less unfairness as compared to complex fairness-aware models, with robust accuracy results (Kung and Yu, 2020). This supports the notion of using interpretable and explainable models for AI in assessments.

- Explainability is concerned about the absence of theoretical basis to justify the development and use of AI system for assessment purposes.

  González-Calatayud, Prendes-Espinosa and Roig-Vila (2021) cites a lack of pedagogical underpinning and AI training, which affects the meaningful development of assessments with pedagogical reference models when AI is applied.

This ethics principle is cited in 42% of the primary studies. This is the joint-second highest cited ethical issue, alongside fairness.

- *Privacy*

  This ethics principle relates to the protection of data subjects against injurious effects from the use of personal information applied in AI systems, without unduly affecting regulatory compliance tied to privacy and restricting AI development.

  - Privacy is concerned about the governance of end-to-end data stewardship, including data collection, storage, disclosure, sharing, security and disposal when applied to AI assessments (Chounta et al., 2022).

    Kiennert et al. (2019) highlights the importance on the management of sensitive data, such as authentication and biometric samples (e.g., data collected for password, voice recognition, facial recognition and/or keystroke detection). Leakage of these data can cause risk of harm. AI systems should be secure and not vulnerable to tampering.

    Stark and Hoey (2021) highlights how individuals are sensitive about data sharing and utilization pertaining to their emotions and emotional expressions. For example, the Facebook emotional contagion study (Kramer, Guillory and Hancock, 2014) were criticized for manipulating emotive content of users.

    Safeguarding of trust and confidence in the governance of data stewardship are important for stakeholder security, privacy and risk of harm (White et al., 2021).

  - Privacy is concerned about the implementation of fair data stewardship practices such as notice, access, and choice (Mougiakou, Papadimitriou and Virvou, 2019).

    Explicit consent should be obtained from data subjects, such that users should be given the right to maintain control over data usage, control the purpose and extent of usage, be granted the option to modify the usage and context, and be given the right to opt in and out of participation. It is noted that consent involving minors can be challenging, as this may require both the students' and their parents' consents (Latham and Goltz, 2019).

    Students who wish to join or withdraw from certain AI-influenced activities may be allowed to do so, especially vulnerable groups such as students with learning disorders, language barriers, or students who come from lower socio-economic backgrounds (Gedrimiene et al., 2020). Educator may also exercise the right to opt in or out of participation in AI-driven teaching evaluation (Tlili

et al., 2018). The possibility to opt in or out may result in data gaps that can affect accuracy of results and research outcomes, indirectly isolate and reveal outcomes of those who opt in or out, and affect discharge of institutional duty to enhance learning experience for students (Tlili et al., 2019).

Merikko et al. (2022) finds that learners are open to sharing data related to demographics and learning performance, but are apprehensive about sharing when it comes to their online behavior, sensitive or process data. Further, the more personal and granular the data are, the less likely the learners will share them. Furthermore, learners who were not performing well are less likely to share their performance data. This may be tied to help-seeking avoidance, as seeking help may be a sign of weakness and a threat to self-esteem (White and Bembenutty, 2013).

- Privacy is concerned about constant surveillance arising from AI use (Megahed, Abdel-Kader and Soliman, 2022).

  There may exist a possibility of violation to individuals' rights to privacy when too much data surveillance exists, especially when data is used beyond academic purposes, for control and surveillance to modify human behavior (Pontual Falcão et al., 2022). Mayfield et al. (2019) discusses the undesirable anxieties and behavioral change related to constant surveillance.

This ethics principle is cited in 55% of the primary studies. This is the highest cited ethical issue.

- *Trust*

  This ethics principle relates to the placing of confidence on (i) AI systems and the (ii) provision of data to achieve assessment objectives. The former is a characteristic of the human-machine relationship formed with an AI system. Low trust is largely linked to the lack of human properties (e.g., lack of affect, emotions, pedagogical intuition) in AI systems (Nazaretsky, Cukurova and Alexandron, 2022). The latter is related to the preservation of privacy.

  - Trust is concerned about the ability to rely on AI systems to make decisions and provide feedback.

    Pontual Falcão et al. (2022) suggests discomfort among students and educators of AI-driven decision making that involves ranking, sorting and classifying individuals, that may reflect political interests, social values, and risks of omissions or biases.

    In addition, the same paper shares that learners and educators were not confident that incorporating AI in assessments can result in improvements of feedback quality.

  - Trust is concerned about the presence of a clear and global consensus of purposes and specifications of AI systems.

For instance, with regards to socio-emotional assessments, Stark and Hoey (2021) highlights a lack of consensus objective agreement on emotion at a global level as an issue. This is because, in the absence of a consensus, the large variation in the implied social and ethical responsibilities have normative implications for AI systems e.g., while considering ethical values such as accountability and fairness when assessing socio-emotional qualities in learners. This affects trust on AI systems.

- Trust is concerned about the autonomy and control that stakeholders have on AI systems.

  For instance, Pontual Falcão et al. (2022) cites educators' discomfort at lack of autonomy and control due to its use in the appraisal of teaching performance and excessive intrusion in learners' learning routine. Learners may also be worried that their autonomy and independent decision making may be deprived.

This ethics principle is cited in 12% of the primary studies.

- *Human centricity*

  This ethics principle relates to the aim towards upholding human agency, dignity and autonomy, minimization of harm (and when necessary, weighed against a greater good), and equitable distribution of benefits.

  - Human centricity is concerned about agency and autonomy of users. This overlaps with the concept of trust.

    Users should not be impacted by profiling, ranking and personalizing derived from AI algorithms. Learning should not be viewed as "*product-oriented learning experiences*" (Duignan, 2020).

    There should be care applied, when it comes to AI algorithm manipulating learner behaviors and emotions (Papa and Jackson, 2021).

    AI systems should not negatively impact a learner’s capacity to learn and his/ her level of autonomy to make learning decisions. There should exist a presence of reversible and clear processes, and the possibility to intervene for blocking, termination, correction and erasure (Mougiakou, Papadimitriou and Virvou, 2019).

  - Human centricity is concerned about fair allocation of resources.

    Users should not have differential access and availability of technology and resources, such that risk of digital divide and technology exclusion arises. Further, the risk of forming unfair groups due to classification and profiling should be mitigated, to ensure equitable learning opportunities (Casas-Roma and Conesa, 2021).

  - Human centricity is concerned about the states of human wellbeing (e.g., psychological wellbeing and satisfaction). A well-designed AI system should seek to achieve positive states of human wellbeing.

This ethics principle is cited in 12% of the primary studies.

- *Cheating*

  This ethics principle relates to dishonest and deceptive learner behavior to violate assessment rules and regulations.

  - Cheating is concerned about the identification of dishonest and deceptive assessment behaviors through the use of AI-based proctoring and plagiarism detection, both in physical venues and remote assessment platforms (Elshafey et al., 2021; Kiennert et al., 2019).

  This ethics principle is cited in 6% of the primary studies.

Paper	AI Ethical Issue(s) as Cited in Paper									
	Inclusivity	Fairness	Accountability	Accuracy	Auditability	Explainability	Privacy	Trust	Human centricity	Cheating
Gupta and Chen (2022)	•	•								
Chounta et al. (2022)		•	•	•	•	•	•			
Deho et al. (2022)		•				•				
Shabaninejad et al. (2022)						•				
Nazaretsky, Cukurova and Alexandron (2022)			•			•		•		
Pontual Falcão et al. (2022)							•	•		
Merikko et al. (2022)							•			
Khairy et al. (2022)				•						
Megahed, Abdel-Kader and Soliman (2022)		•					•			
Conati et al. (2021)						•				
González-Calatayud, Prendes-Espinosa and Roig-Vila (2021)						•				
White et al. (2021)							•			
Ahn et al. (2021)				•						
Stark and Hoey (2021)		•	•				•	•		
Papa and Jackson (2021)		•				•			•	
Kim et al. (2021)				•						
Litman et al. (2021)		•		•		•				
Casas-Roma and Conesa (2021)		•			•	•	•		•	
Costas-Jauregui et al. (2021)	•		•				•	•		
Elshafey et al. (2021)										•
Schneider, Dowell and Thompson (2021)	•			•		•				

Gedrimiene et al. (2020)	•		•				•			
Kumar and Boulanger (2020)						•				
Khosravi, Sadiq and Gasevic (2020)			•				•			
Martín Núñez and Lantada (2020)	•						•			
Hakami and Hernández-Leo (2020)		•	•			•			•	
Mougiakou, Papadimitriou and Virvou (2019)							•		•	
Mayfield et al. (2019)		•		•			•			
Latham and Goltz (2019)		•	•			•	•			
Tlili et al. (2019)	•	•	•		•	•	•			
Kiennert et al. (2019)				•			•			•
Peña-Ayala (2018)		•					•			
Tlili et al. (2018)		•	•	•			•			
Total Count	**6**	**14**	**10**	**9**	**3**	**14**	**18**	**4**	**4**	**2**
Percentage	*18%*	*42%*	*30%*	*27%*	*9%*	*42%*	*55%*	*12%*	*12%*	*6%*

Table 3. Breakdown of sub-themes of ethical issues by paper

4.5 Second pass of topic modelling and network analyses

Next, we utilize the keyword corpuses of fourteen sub-themes of AI application areas and ten sub-themes of ethical issues as an input, to perform the second pass of topic modelling, and network analyses.

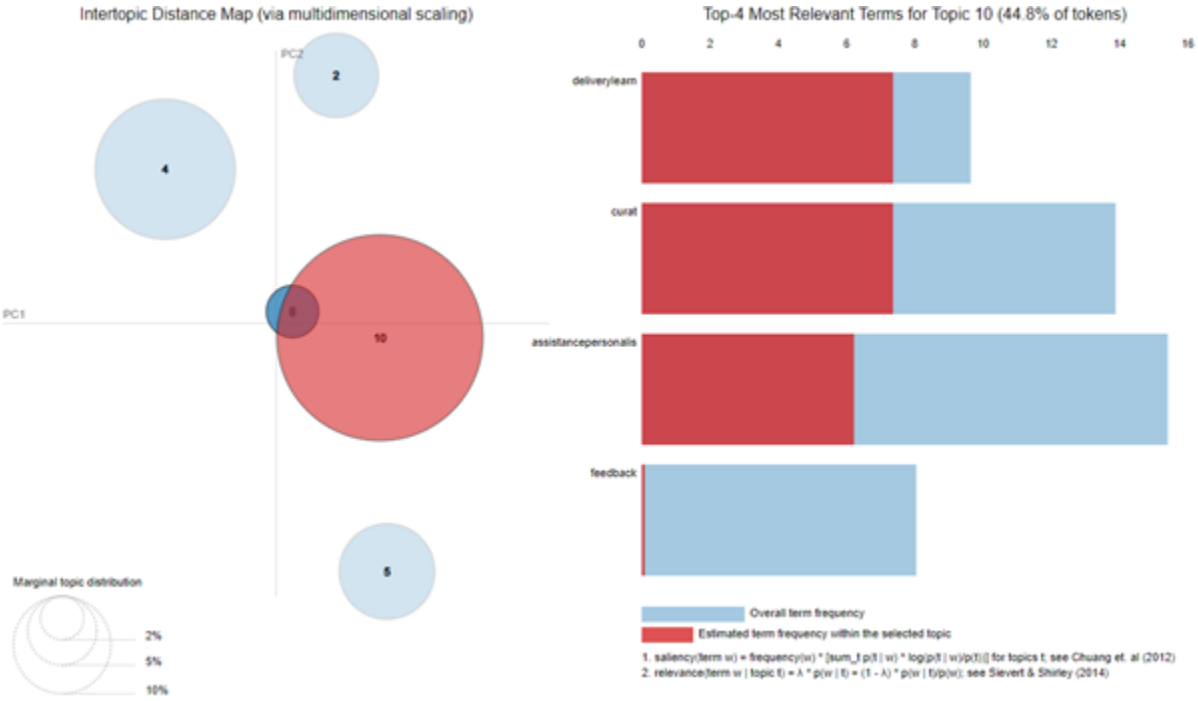

Figure 9. Topic modelling of corpuses involving AI application areas and related ethical principles

Latent Topic	Percentage of Tokens	Top Keywords
System design and check	*7.5%*	System; Design; Review
Data stewardship and surveillance	*2.9%*	Privacy; Sensitive; Data
Assessment construction and rollout	*44.8%*	Deliver; Curate; Personalize
Assessment administration	*9.6%*	Proctor; Plagiarism; Cheat
Grading and evaluation	*20.7%*	Evaluation; Feedback; Response

Table 4. Second pass of topic modelling – Latent topic and top keywords

Figure 10. Network analyses of corpuses involving AI application areas and related ethical principles

Research identified five topical archetypes via topic modelling. For instance, in Fig. 9, we observed the dominant latent topic linked to AI-based assessment construction and rollout aspects. This mirrored well with the network analyses visualization in Fig. 10. In the network analysis diagram, we observed a clear clustering of five topics, with *Assessment Construction and Rollout* similarly dominant in the cluster diagram. The top keywords and latent topics of topic modeling are shown in Table 4.

Understanding of these five key archetypical themes allow researchers and practitioners to breakdown the landscape into clear segments, to decide which area they would like to develop further insights and applications.

DISCUSSION

Morley et al. (2020) emphasizes on the importance of the translation of AI principles into the 'what' and 'how' of implementation. Building on the work of preceding sections, this section discusses actionable insights to make the addressing of AI ethics operable in the real world.

Research Question (RQ)4: What are the key themes of the systematic literature map?

Ontology can be defined as "*an explicit specification of a conceptualization*" (Gruber, 1993), geared towards a "*shared taxonomy of entities*" (Smith and Welty, 2001), as applied in information systems. This is opposed to the philosophical use of the concept of ontology as a nature of reality (Ashok et al., 2022). The investigation of AI systems can be considered a sub-field of information systems.

Ashok et al., (2022) describes three fundamental domains to conceptually represent the interweaving ethical elements and interrelationships inherent in the design and application of AI in digital technologies. This is theoretically underpinned by ontological frameworks of Ogden and Richards (1923), Popper (1979) and Project and Peirce (1998).

This triadic framework is a modular architecture of an assemblage of technological components that consist of the:

- *Physical domain* (or the referent or object in semiotics): This includes the: (i) *device layer* which comprises a logical capability operating system layer, and the physical machinery hardware layer; and the (ii) *network layer* which comprises the logical transmission network protocol layer and physical network transport layer.

  Some relevant applications are author systems, intelligent tutoring shells, AI-integrated learning environments, educational robotics, and AI collaborative tools.

- *Cognitive domain* (or the symbol or science in semiotics): This comprises the *content layer* where data is stored, created, mapped, manipulated, utilized, and shared.

  Some relevant examples are multimodal structured contents of text, and unstructured contents of images, sounds and videos of assessment submissions. This layer also provides the metadata and directory information of users, content tags, location stamps, time stamps, encoding and copyright etc.

- *Information domain* (or the reference or interpretant in semiotics): This comprises the *service layer* which encompasses the functionality of the application and its interaction with users, underpinned by AI algorithms.

  Some relevant examples are use of knowledge representation for instructions, human factor and interface design, and AI-integrated visualization and graphics for feedbacks.

We extend the triadic ontological framework as described by Ashok et al., (2022) to model and visualize the systematic literature map of this chapter (Fig. 11). We note that, over and above the triadic domains, Ashok et al., (2022) further describes a governance domain, defined by Floridi, (2018) as *"the practice of establishing and implementing policies, procedures, and standards for the proper development, use, and management of the infosphere*." In our opinion, governance is a key consideration across all triadic domains, and hence, the governance domain is not explicitly illustrated in our framework.

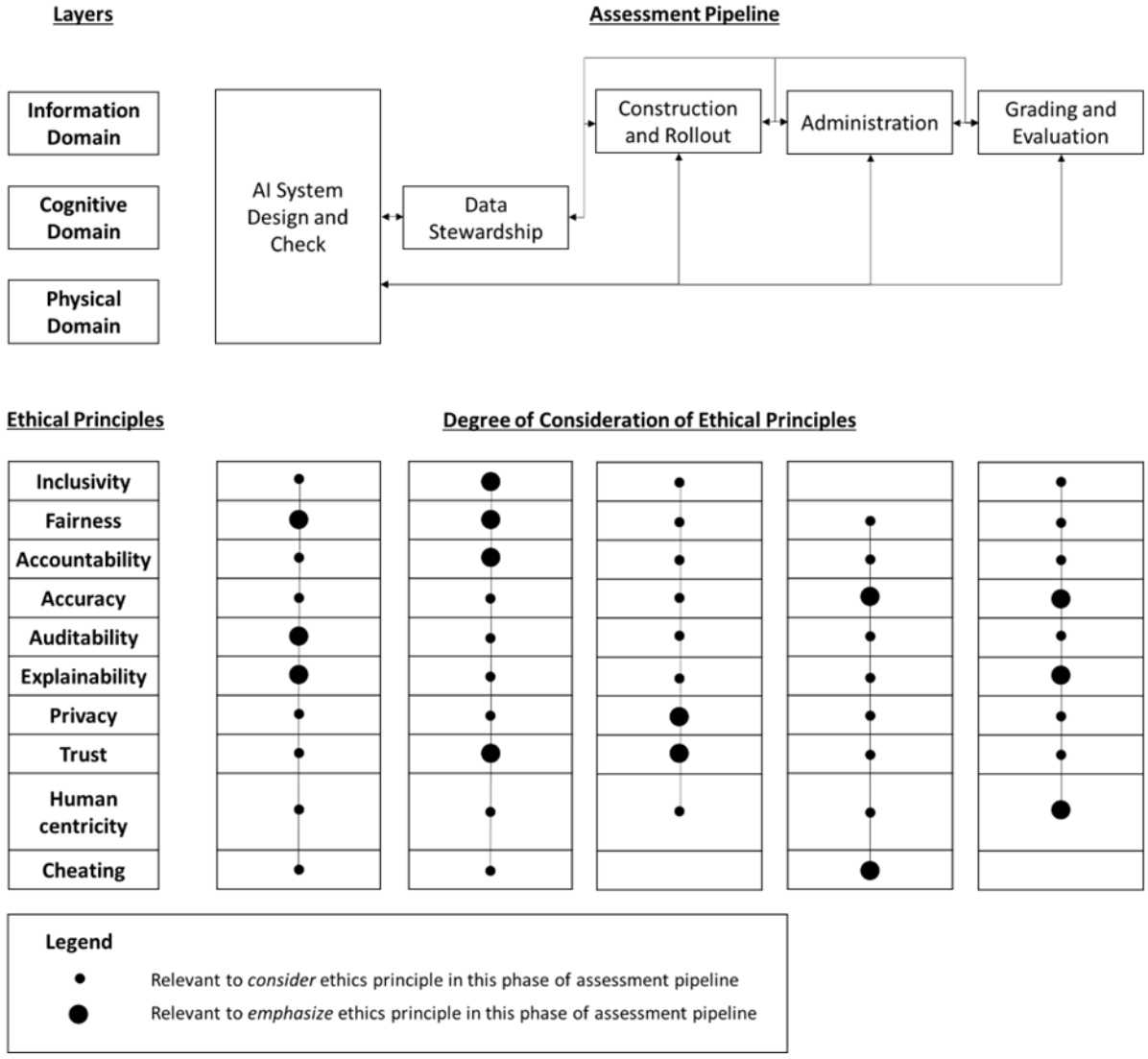

Figure 11. Visualization of the systematic literature map of key research themes

The five distinct clusters identified by topic modelling and network analyses in Fig. 9 and Fig. 10 are mapped to the triadic ontological framework in Fig. 11, as follows:

- *AI system design and check for assessment purposes*

  This phase extends across the physical, cognitive and information domains, and is involved with the design, implementation and maintenance of the AI system for system interactivity, robustness and security. From a predictive analytics point of view, the model constructed should be appropriate – upholding accuracy, inclusivity, accountability, privacy, trust and human centricity.

At this phase, the overriding ethics considerations are explainability and auditability. The AI system should be created with clear, easy-to-understand and transparent protocols, so that relevant stakeholders and independent third-party auditors can review the processes, perform interventions, mitigate issues, and enable redress in an event of negative outcomes that may arise. In addition, fairness is concerned about the treatment of algorithmic bias to ensure diversity, equity, non-prejudice and non-favoritism towards learners' sensitive attributes, so that needs of minority groups are not disadvantaged or underrepresented.

- *Data stewardship and surveillance*

  This phase extends across the cognitive and information domains, and is involved with the governance and implementation of good data stewardship, and appropriate surveillance practices (if any).

  At this phase, the overriding ethics consideration is privacy. One instance is behavioral surveillance, which may be a violation to human rights to privacy especially when data is used beyond academic purposes, for control and surveillance to modify human behavior. In addition, trust is also an important facet concerned about the preservation of privacy when sensitive data are disclosed.

- *AI-based assessment construction and rollout*

  This phase is predominantly situated in the information domain, and is involved with the construction, curation or delivery of assessment, the communication of evaluation and feedback with stakeholders via AI-integrated communication dashboards, and the carrying out of interventions and assistances to improve assessment and evaluation performance. Assessment and evaluation can be in the form of formative (or summative) individual (or group) cognitive (or socio-emotional) assessment. It can also be a form of teaching evaluation.

  At this phase, the overriding ethics considerations are inclusivity and fairness, so that appropriate and equitable assessments and evaluations are rolled out, embracing diversity, empathy and sensitivity towards the evaluated stakeholders. Furthermore, accountability is an important ethics consideration, as there should exist a responsible discharge of AI ethical principles and compliance with relevant rules and guidelines, when designing and delivering AI-driven assessments. In addition, there should exist trust and confidence on AI systems to achieve assessment and evaluation objectives.

- *Administration of assessments using AI systems*

  This phase is predominantly situated in the information domain, and is involved with the administration of assessment and evaluation, which may comprise authentication and security measures, proctoring and/or plagiarism detection.

  At this phase, the overriding ethics considerations are the overcoming of cheating violations, and the application of accuracy to correctly identify assessment candidates and cheating cases.

- *AI-facilitated assessment grading and evaluation*

This phase is predominantly in the information domain, and is involved primarily with the interpretation of textual and/or audio-visual responses collected by AI systems, the evaluation of performance, and the provision of feedbacks. These may be performed by autonomous intelligent agents. From an educator's point of view, this phase may involve the evaluation of teaching effectiveness.

At this phase, the overriding ethics considerations are explainability, so evaluators can understand and adjudge if the grading and/or ranking is accurate and reliable. In addition, there is an element of human centricity. This largely relates to the agency and autonomy of human users, in the presence of AI-generated decisions, and the capacity to intervene for correction and redress.

There is an emphasis that the framework does not draw clear delineations when categorizing AI assessment use cases across triadic domains. For instance, the *Grading and Evaluation* research theme is predominantly arising from the cognitive domain. However, coding and rolling out a moral reasoning AI system for AI-generated decisions, evaluations, responses and feedbacks, a sub-item of this research theme, may straddle across all cognitive, information and physical domains. This said, the framework provides a guide to generalize observed phenomena.

Theoretical implications

Understanding the ethical dimensions of the application of AI in assessments across the assessment creation pipeline, in a conceptual framework as shown in Fig. 11, are but the first steps toward informing a critical awareness and a more holistic approach in the governance, stewardship and regulation practices of ethics in this subject matter.

To steer this ethical discourse, building upon the framework proposed by Floridi (2018), it is useful to consider what are (i) politically feasible, (ii) culturally sustainable, (iii) socially acceptable, (iv) institutionally preferred, and (v) legally or rule enforceable, before elaborating and enforcing a system of rules to regulate agents' ethical behaviors. This requires a proactive and constructive normative cascade underpinned by theoretical and empirical foundations, rather than a reactive add-on afterthought that arise from societal backlashes, e.g., when stakeholders are negatively affected or when the public rejects legal recommendations.

A significant aspect of the theoretical discourse in this study discusses about hard ethics (i.e., the morally good or bad, or right or wrong duties, behavior, responsibilities, values and rights). This is a fundamentally important discourse as hard ethics helps make and shape regulations. However, there also exists the plane of soft ethics, that embodies normative ground over and above hard ethics, as a "*post-compliance ethics*" (Floridi, 2018). Even if regulations may already exist in the right side of the moral against the immoral divide, regulations do not cover everything, and human agents will need to leverage on ethics evaluation to guide and perform self-regulation of morality, especially if competing interests or values arise that need to be assessed and adjudged. Only then can we achieve good moral citizenry towards a mature infosphere.

Practical implications

The triadic framework as discussed in this chapter provides a thematic guide to future applied research. In the translation of principles to practice, there may exist ethical risks that should be avoided. Floridi (2021) highlights the following "*ethics regunds*", namely: (i) *ethics shopping*, or the picking and choosing of ethics principles that are justified as a posteriori and retrofitted to pre-existing behaviors, (ii) *ethics bluewashing*, or the implementation of superficial or misleading measures to appear ethical, (iii) *ethics lobbying*, or the use of ethics to avoid or delay good and necessary regulation and enforcement, (iv) *ethics dumping*, or the export or import of unethical activities to a place with less strict regulations, and (v) *ethics shirking*, or the engaging of less ethical works over a period of time to lower the perceived resistance against such works. A misjudgment or misunderstanding can over time, in the lens of Socrates, lead to ethical malpractices.

CONCLUSION

Ethical check and balances should be put in place with the increasingly pervasive use of AI, especially when its growth trajectory appears seemingly aimed towards anthropomorphism, a reality further taking shape by recent advances in generative AI. Possible violations of fundamental human ethics in a societal institution as important as education should be looked upon with scrutiny.

In this study, we looked at how the design and use of AI in education, and in particular, assessments, can conform as closely as possible to basic ethical principles. We systematically investigated the key assessment components and ethical principles highlighted in existing literature, mapped them across the end-to-end assessment pipeline while accounting for different assessment types, and constructed a systematic literature mapping framework highlighting key archetypical research themes. The proposed systematic literature mapping framework allows researchers and practitioners to deep dive into key thematic research areas, facilitating practical implementation of ethics programs and activities in educational institutions.

Research identified five key archetypical research themes, namely (i) AI system design and check for assessment purposes, (ii) data stewardship and surveillance, (iii) AI-based assessment construction and rollout, (iv) administration of assessments using AI systems, and (v) AI-facilitated assessment grading and evaluation. Ten literature-derived ethical principles, namely, accuracy, privacy, human centricity, fairness, inclusivity, trust, explainability, cheating, accountability and auditability, were mapped to these research themes. The study summarizes and rationalizes the impact of each ethical element in each research area, and discusses theoretical and practical implications of the findings.

As highlighted in the limitations section, future work can extend the use of literature databank beyond Scopus, to include e.g., Web of Science, IEEE Xplore or EBSCO Host, in the systematic literature mapping exercise. Furthermore, to account for thematic diversity, the intrinsic value of the diversification of the archetypical research themes can be studied.

While this study is based upon the subject of assessments, the ethical elements of the discourse has relevance beyond assessments, and can be applied to other areas of AIED. Other future works can contribute to the examination on the underpinning theories relating the ontological, semantics, and the epistemological deliberations and practical applications of ethics in this subject matter, across the spheres of philosophy, learning, psychology, sociology and

technology. In addition, practical applications of the actionable insights in this chapter, in a form of strategic and operational frameworks or case studies, can be another pragmatic endeavor by practitioners and researchers.

Herwix et al. (2022) highlighted the importance of more serious and systematic engagement with the selection, framing and prioritization of ethical issues. There is an emphasis among the state-of-the-art for the need to be more aware, anticipatory, reflecting and informed about the variety of perspectives and contemporary debates concerning AIED ethics. In particular, the relevancy and idiosyncrasy to assessments in our study can help bring forward distinctive actionable applications in this realm.

REFERENCES

Abdi, S., Khosravi, H., Sadiq, S., & Gasevic, D. (2020, March). Complementing educational recommender systems with open learner models. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 360-365). ACM.

Ahn, J., Nguyen, H., Campos, F., & Young, W. (2021, April). Transforming everyday information into practical analytics with crowdsourced assessment tasks. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 66-76). ACM.

Andersson, H., Svensson, A., Frank, C., Rantala, A., Holmberg, M., & Bremer, A. (2022). Ethics education to support ethical competence learning in healthcare: An integrative systematic review. *BMC Medical Ethics, 23,* 29. DOI: https://doi.org/10.1186/s12910-022-00766-z.

Andrews, D., Leitner, P., Schön, S., & Ebner, M. (2022). Developing a Professional Profile of a Digital Ethics Officer in an Educational Technology Unit in Higher Education. In *International Conference on Human-Computer Interaction* (pp. 157-175). Springer, Cham.

Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for artificial intelligence and digital technologies. *International Journal of Information Management*, *62*, 102433.

Australian Government. (2019). Australia's Artificial Intelligence Ethics Framework. Department of Industry, Science and Resources. [Online]. Canberra, Australia. Retrieved: https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework [Accessed 20 Nov 2022].

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... & Rahwan, I. (2018). The moral machine experiment. *Nature*, *563*(7729), 59-64.

Aynaud, T. (2020). Python-louvain. Louvain algorithm for community detection. [Online]. Retrieved: https://github.com/taynaud/python-louvain [Assessed 15 November 2022].

Bastani, H., Bastani, O., & Kim, C. (2018). Interpreting predictive models for human-in-the-loop analytics. *arXiv preprint arXiv:1705.08504*.

Bessen, J., Impink, S. M., & Seamans, R. (2022, July). The cost of ethical AI development for AI startups. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 92-106). ACM.

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, *181*, 21-34.

Blanchard, E. G. (2012). On the weird nature of ITS/AIED conferences. In *International Conference on Intelligent Tutoring Systems* (Vol. *7315*, pp. 280–285). Springer.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

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

Bozkurt, A. (2020). Educational technology research patterns in the realm of the digital knowledge age. *Journal of Interactive Media in Education*, *2020*(1).

Breucker, P., Cointet, J., Hannud Abdo, A., Orsal, G., de Quatrebarbes, C., Duong, T., Martinez, C., Ospina Delgado, J. P., Medina Zuluaga, L. D., Gómez Peña, D. F., Sánchez Castaño, T. A., Marques da Costa, J., Laglil, H., Villard, L., & Barbier, M. (2016). CorTexT Manager (version v2). [Online]. Retrieved: https://docs.cortext.net [Accessed 24 Oct 2022].

Campedelli, G. M. (2021) Where are we? Using Scopus to map the literature at the intersection between artificial intelligence and research on crime. *Journal of Computational Social Science, **4**,* 503–530. DOI: https://doi.org/10.1007/s42001-020-00082-9.

Casas-Roma, J., & Conesa, J. (2021). A literature review on artificial intelligence and ethics in online learning. *Intelligent Systems and Learning Data Analytics in Online Education*, 111-131. DOI: https://doi.org/10.1016/B978-0-12-823410-5.00006-1.

Chassang, G., Thomsen, M., Rumeau, P., Sèdes, F., & Delfin, A. (2021). An interdisciplinary conceptual study of Artificial Intelligence (AI) for helping benefit-risk assessment practices. *AI Communications, 3*(4), 1-26.

Chaudhry, M. A., & Kazim, E. (2022). Artificial Intelligence in Education (AIEd): a high-level academic and industry note 2021. *AI and Ethics*, *2*(1), 157-165.

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *1*, 100002.

Chounta, I. A., Bardone, E., Raudsep, A., & Pedaste, M. (2022). Exploring teachers' perceptions of artificial intelligence as a tool to support their practice in Estonian K-12 education. *International Journal of Artificial Intelligence in Education*, *32*(3), 725-755.

Conati, C., Barral, O., Putnam, V., & Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial Intelligence*, *298*, 103503.

Costas-Jauregui, V., Oyelere, S. S., Caussin-Torrez, B., Barros-Gavilanes, G., Agbo, F. J., Toivonen, T., Motz, R., & Tenesaca, J. B. (2021, October). Descriptive analytics dashboard for an inclusive learning environment. In *2021 IEEE Frontiers in Education Conference (FIE)* (pp. 1-9). IEEE.

Deho, O. B., Zhan, C., Li, J., Liu, J., Liu, L., & Duy Le, T. (2022). How do the existing fairness metrics and unfairness mitigation algorithms contribute to ethical learning analytics? *British Journal of Educational Technology, 53,* 822–843.

Duignan, P. A. (2020). Navigating the future of learning: the role of smart technologies. *Leading Educational Systems and Schools in Times of Disruption and Exponential Change: A Call for Courage, Commitment and Collaboration,* pp. 125–137. Emerald Publishing Limited, Bingley. DOI: https://doi.org/10.1108/978-1-83909-850-520201012/full/html.

Eicher, B., Polepeddi, L., & Goel, A. (2018). Jill Watson doesn't care if you're pregnant: Grounding AI ethics in empirical studies. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 88-94). ACM.

Elsevier (2022). How Scopus works, Scopus contents. [Online]. Retrieved: https://www. elsevier.com/solutions/scopus/how-scopus-works/content [Assessed 15 November 2022].

Elshafey, A. E., Anany, M. R., Mohamed, A. S., Sakr, N., & Aly, S. G. (2021). Dr. Proctor: A multi-modal AI-based platform for remote proctoring in education. In *International Conference on Artificial Intelligence in Education* (pp. 145-150). Springer, Cham.

European Parliament (2021). Report on artificial intelligence in education, culture and the audiovisual sector (2020/2017(INI)). Committee on Culture and Education. [Online]. Retrieved: https://www.europarl.europa.eu/ doceo/document/A-9-2021-0127_EN.html [Accessed 20 Nov 2022].

Fahimnia, B., Sarkis, J., & Davarzani, H. (2015). Green supply chain management: A review and bibliometric analysis. *International Journal of Production Economics*, *162*, 101–114. DOI: https://doi.org/10.1016/j.ijpe.2015.01.003.

Floridi, L. (2018). Soft ethics and the governance of the digital. *Philosophy & Technology*, *31*(1), 1-8.

Floridi, L. (2021). Translating principles into practices of digital ethics: Five risks of being unethical. In *Ethics, Governance, and Policies in Artificial Intelligence* (pp. 81-90). Springer, Cham.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1).

Gedrimiene, E., Silvola, A., Pursiainen, J., Rusanen, J., & Muukkonen, H. (2020). Learning analytics in education: Literature review and case examples from vocational education. *Scandinavian Journal of Educational Research*, *64*(7), 1105-1119.

González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. *Applied Sciences*, *11*(12), 5467.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199–220.

Gupta, S., & Chen, Y. (2022). Supporting inclusive learning using chatbots? A chatbot-led interview study. *Journal of Information Systems Education*, *33*(1), 98-108.

Hakami, E., & Hernández Leo, D. (2020). How are learning analytics considering the societal values of fairness, accountability, transparency and human well-being: A literature review. In Martínez-Monés A, Álvarez A, Caeiro-Rodríguez M, Dimitriadis Y (Eds.)*, Learning Analytics Summer Institute Spain 2020* (pp. 121-41). Aachen: CEUR. Valladolid, Spain.

Hendrycks, D., Carlini, N., Schulman, J., & Steinhardt, J. (2021). Unsolved problems in ML safety. *arXiv preprint arXiv:2109.13916*.

Hanna, R., & Kazim, E. (2021). Philosophical foundations for digital ethics and AI ethics: a dignitarian approach. *AI and Ethics*, *1*(4), 405-423.

Heo, J., & Lee, J. (2019, July). CiSA: An inclusive chatbot service for international students and academics. In *International Conference on Human-Computer Interaction* (pp. 153-167). Springer, Cham.

Herwix, A., Haj-Bolouri, A., Rossi, M., Tremblay, M. C., Purao, S., & Gregor, S. (2022). Ethics in information systems and design science research: Five perspectives. *Communications of the Association for Information Systems*, *50*(1), 589-616.

High-Level Expert Group on Artificial Intelligence (AI HLEG). (2019). *Ethics guidelines for Trustworthy AI.* European Commission. [Online]. Retrieved https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html [Accessed 22 Nov 2022].

Hill, P., and Barber, M. (2014). *Preparing for a renaissance in assessment.* Pearson, London.

Hoekstra, A., & Kaptein, M. (2021). The integrity of integrity programs: Toward a normative framework. *Public Integrity*, *23*(2), 129-141.

Hu, Q., & Rangwala, H. (2020). Towards fair educational data mining: A case study on detecting at-risk students. In *Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020)* (pp. 431–437). ERIC.

Hussar, W.J., and Bailey, T.M. (2020). *Projections of Education Statistics to 2028* (NCES 2020-024). U.S. Department of Education, Washington, DC: National Center for Education Statistics.

Hwang, G. J., & Chien, S. Y. (2022). Definition, roles, and potential research issues of the metaverse in education: An artificial intelligence perspective. *Computers and Education: Artificial Intelligence*, 100082.

IEEE. (2019). Ethically aligned design. [Online]. Retrieved: https://ethicsinaction.ieee.org/ [Accessed 20 Nov 2022].

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence*, *2*, 100017.

Khairy, D., Alkhalaf, S., Areed, M. F., Amasha, M. A., & Abougalala, R.A. (2022). An algorithm for providing adaptive behavior to humanoid robot in oral assessment. *International Journal of Advanced Computer Science and Applications, 13*(9). DOI: http://dx.doi.org/10.14569/IJACSA.2022.01309119.

Khan, A. A., Badshah, S., Liang, P., Waseem, M., Khan, B., Ahmad, A., ... & Akbar, M. A. (2022, June). Ethics of AI: A systematic literature review of principles and challenges. In *Proceedings of the International Conference on Evaluation and Assessment in Software Engineering 2022* (pp. 383-392).

Khosravi, H., Demartini, G., Sadiq, S., & Gasevic, D. (2021, April). Charting the design and analytics agenda of learnersourcing systems. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 32-42). ACM.

Khosravi, H., Sadiq, S., & Gasevic, D. (2020, February). Development and adoption of an adaptive learning system: Reflections and lessons learned. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (pp. 58-64). ACM.

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J., ... & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *3*, 100074.

Kiennert, C., De Vos, N., Knockaert, M., & Garcia-Alfaro, J. (2019). The influence of conception paradigms on data protection in e-learning platforms: A case study. *IEEE Access*, *7*, 64110-64119.

Kim, J. H., Baek, J., Hwang, C., Bae, C., & Park, J. (2021, June). Condensed discriminative question set for reliable exam score prediction. In *International Conference on Artificial Intelligence in Education* (pp. 446-450). Springer, Cham.

Klinger, J., Mateos-Garcia, J., & Stathoulopoulos, K. (2020). A narrowing of AI research? *arXiv preprint arXiv:2009.10385*.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788-8790.

Kumar, V., & Boulanger, D. (2020, October). Explainable automated essay scoring: Deep learning really has pedagogical value. In *Frontiers in Education* (Vol. 5, p. 572367). Frontiers Media SA.

Kung, C., & Yu, R. (2020). Interpretable models do not compromise accuracy or fairness in predicting college success. In *Proceedings of the Seventh ACM Conference on Learning@ Scale* (pp. 413–416). ACM.

Last, M., & Danon, G. (2020). Automatic question generation. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(6), e1382.

Latham, A., & Goltz, S. (2019, June). A survey of the general public's views on the ethics of using AI in education. In *International Conference on Artificial Intelligence in Education* (pp. 194-206). Springer, Cham.

Leslie, D. (2019). Understanding artificial intelligence ethics and safety. *arXiv preprint arXiv:1906.05684*.

Lim, T., Gottipati, S, & Cheong, M. (2022). Authentic Assessments for Digital Education: Learning Technologies Shaping Assessment Practices. *Proceedings of the 30th International Conference on Computers in Education (ICCE 2022).* 1, p. 587-592. Kuala Lumpur, Malaysia. ISBN: 978-986-972-149-3.

Litman, D., Zhang, H., Correnti, R., Matsumura, L. C., & Wang, E. (2021, June). A fairness evaluation of automated methods for scoring text evidence usage in writing. In *International Conference on Artificial Intelligence in Education* (pp. 255-267). Springer, Cham.

Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour, 1*, 0028. DOI: https://doi.org/10.1038/s41562-016-0028.

Martin Nunez, J. L., & Diaz Lantada, A. (2020). Artificial intelligence aided engineering education: State of the art, potentials and challenges. *International Journal of Engineering Education*, *36*(6), 1740-1751.

Martinez-Maldonado, R., Gaševic, D., Echeverria, V., Fernandez Nieto, G., Swiecki, Z., & Buckingham Shum, S. (2021). What Do You Mean by Collaboration Analytics? A Conceptual Model. *Journal of Learning Analytics*, *8*(1), 126-153.

Mayfield, E., Madaio, M., Prabhumoye, S., Gerritsen, D., McLaughlin, B., Dixon-Román, E., & Black, A. W. (2019, August). Equity beyond bias in language technologies for education. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 444-460). Association for Computational Linguistics.

Megahed, N. A., Abdel-Kader, R. F., & Soliman, H. Y. (2022). Post-pandemic education strategy: Framework for artificial intelligence-empowered education in engineering (AIEd-Eng) for lifelong learning. In *International Conference on Advanced Machine Learning Technologies and Applications* (pp. 544-556). Springer, Cham.

Merikko, J., Ng, K., Saqr, M., & Ihantola, P. (2022). To opt in or to opt out? Predicting student preference for learning analytics-based formative feedback. *IEEE Access*, *10*, 99195-99204.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. DOI: https://doi.org/10.1371/journal.pmed.1000097.

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics, 26*(4), 2141–2168.

Mougiakou, E., Papadimitriou, S., & Virvou, M. (2018, July). Intelligent tutoring systems and transparency: The case of children and adolescents. In *2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA)* (pp. 1-8). IEEE.

Nazaretsky, T., Cukurova, M., & Alexandron, G. (2022). An instrument for measuring teachers' trust in AI-based educational technology. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 56-66). ACM.

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. P. T. (2022). Ethical principles for artificial intelligence in education. *Education and Information Technologies*. DOI: https://doi.org/10.1007/s10639-022-11316-w.

Ogden, C. R., & Richards, I. A. (1923). *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism.* London Routledge & Kegan Paul.

Papa, R., & Jackson, K. M. (2021). Enduring questions, innovative technologies: Educational theories interface with AI. In *Intelligent Computing* (pp. 725-742). Springer, Cham.

Peña-Ayala, A. (2018). Learning analytics: A glance of evolution, status, and trends according to a proposed taxonomy. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(3), e1243.

Pereira, J. (2016). Leveraging chatbots to improve self-guided learning through conversational quizzes. *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 911-918). ACM.

Petersen, K., & Gencel, C. (2013). Worldviews, research methods, and their relationship to validity in empirical software engineering research. In *2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement* (pp. 81-89). IEEE.

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, *64*, 1-18.

Popper, K. (1979). *Three worlds*. University of Michigan.

Pontual Falcão, T., Lins Rodrigues, R., Cechinel, C., Dermeval, D., Harada Teixeira de Oliveira, E., Gasparini, I., ... & Ferreira Mello, R. (2022, March). A Penny for your Thoughts: Students and Instructors' Expectations about Learning Analytics in Brazil. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 186-196).

Project, P. E., & Peirce, C. S. (1998). *The essential peirce* (Volume 2). Indiana University Press.

Raji, I. D., Scheuerman, M. K., & Amironesei, R. (2021, March). You can't sit with us: Exclusionary pedagogy in ai ethics education. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 515-525). ACM.

Rodrigues, S. P., van Eck, N. J., Waltman, L., & Jansen, F. W. (2014). Mapping patient safety: A large-scale literature review using bibliometric visualisation techniques. *BMJ Open*, *4*(3), e004468. DOI: https://doi.org/10.1136/bmjopen-2013-004468.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1355.

Sadler, D. R. (1989). Formative assessment in the design of instructional systems. *Instructional Science 18*, 119–144.

Sánchez-Prieto, J. C., Gamazo, A., Cruz-Benito, J., Therón, R., & García-Peñalvo, F. J. (2020). AI-driven assessment of students: Current uses and research trends. In *P. Zaphiris & A. Ioannou* (Eds.), Learning and Collaboration Technologies. Design, Experiences. 7th International Conference, LCT 2020, Held as part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I (pp. 292-302). Springer Nature. DOI: https://doi.org/10.1007/978-3-030-50513-4_22.

Schiff, D. (2021). Out of the laboratory and into the classroom: the future of artificial intelligence in education. *AI & society*, *36*(1), 331-348.

Schneider, B., Dowell, N., & Thompson, K. (2021). Collaboration analytics—current state and potential futures. *Journal of Learning Analytics*, *8*(1), 1-12.

Shabaninejad, S., Khosravi, H., Abdi, S., Indulska, M., & Sadiq, S. (2022, June). Incorporating explainable learning analytics to assist educators with identifying students in need of attention. In *Proceedings of the Ninth ACM Conference on Learning@ Scale* (pp. 384-388). ACM.

Shapiro, J., & Blackman, R. (2020). Four steps for drafting an ethical data practices blueprint. *TechCrunch.* [Online]. Retrieved*: https://techcrunch.com/2020/07/24/four-steps-for-an-ethical-data-practices-blueprint/* [Assessed 15 November 2022].

Sibai, F. N. (2020, June). AI crimes: a classification. In *2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)* (pp. 1-8). IEEE.

Sievert, C., & Shirley, KE. (2014). LDAvis: A method for visualizing and interpreting topics, *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (pp. 63–70). Baltimore, Maryland, USA.

Siau, K., & Wang, W. (2020). Artificial intelligence (AI) ethics: ethics of AI and ethical AI. *Journal of Database Management*, *31*(2), 74-87.

Smith, B., & Welty, C. (2001). Ontology: Towards a new synthesis. *Formal Ontology in Information Systems*. ACM Press.

Stahl, B. C., Timmermans, J. O. B., & Mittelstadt, B. D. (2016). The ethics of computing: A survey of the computing-oriented literature. *ACM Computing Surveys, 48*(4), 1–38.

Stark, L., & Hoey, J. (2021). The ethics of emotion in artificial intelligence systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 782-793). ACM.

Stringer, R. (2018). Realist ethical naturalism for ethical non-naturalists. *Philosophical Studies*, *175*(2), 339-362.

The Open University. (2014). *Policy on Ethical use of Student Data for Learning Analytics*. [Online]. Retrieved: https://help.open.ac.uk/documents/policies/ethical-use-of-student-data/files/22/ethical-use-of-student-data-policy.pdf [Accessed: 15 Nov 2022].

Tlili, A., Essalmi, F., Jemni, M., & Chen, N. S. (2018). A complete validated learning analytics framework: Designing issues from data preparation perspective. *International Journal of Information and Communication Technology Education (IJICTE)*, *14*(2), 1-16.

Tlili, A., Essalmi, F., Jemni, M., & Chen, N. S. (2019). A complete validated learning analytics framework: designing issues from data use perspective. *International Journal of Information and Communication Technology Education (IJICTE)*, *15*(3), 42-59.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2021). Recommendation on the ethics of artificial intelligence. UNESDOC Digital Library. [Online]. Retrieved: https://unesdoc.unesco.org/ark:/48223/pf0000380455 [Accessed 20 Nov 2022].

van Otterlo, M. (2017). From algorithmic black boxes to adaptive white boxes: Declarative decision-theoretic ethical programs as codes of ethics. *arXiv preprint arXiv:1711.06035*.

Weitzman, M. L. (1993). What to preserve? An application of diversity theory to crane conservation. *The Quarterly Journal of Economics*, *108*(1), 157-183.

White, M. C., & Bembenutty, H. (2013). Not all avoidance help seekers are created equal: Individual differences in adaptive and executive help seeking. *Sage Open*, *3*(2), 2158244013484916.

White, J. P., Dennis, S., Tomko, M., Bell, J., & Winter, S. (2021). Paths to social licence for tracking-data analytics in university research and services. *PloS one*, *16*(5), e0251964.

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). *Ethical and societal implications of algorithms, data, and artificial intelligence: A roadmap for research*. London: Nuffield Foundation.

Williamson, B. (2021). Psychodata: disassembling the psychological, economic, and statistical infrastructure of 'social-emotional learning.' *Journal of Education Policy, 36*(1), 129-154.

KEY TERMS AND DEFINITIONS

Accountability: This ethics principle relates to the responsible discharge of AI ethics when designing and delivering AI systems, depending on the roles and contexts, in a consistent manner.

Accuracy: In the context of educational assessments, this ethics principle relates to the reliability and validity of assessments when an AI system is applied.

Applied ethics: The study of the practical application of philosophical tools to examine and provide solutions to real world morality issues.

Artificial intelligence in education (AIED): The machine mimicry of human-like consciousness and behavior to achieve educational goals, through the use of technology that allows digital systems to perform tasks commonly associated with intelligent beings.

Auditability: This ethics principle relates to the permitting of independent third-party reviewers to audit, analyze and report findings relating to the usage and design of data and AI algorithms in education.

Cheating: This ethics principle relates to dishonest and deceptive learner behavior to violate educational rules and regulations.

Consequentialism: A type of normative ethics that emphasizes that the outcome of an action defines the morality of an action.

Deontological ethics: A type of normative ethics that emphasizes on an individual's rights and duties, including the presence of natural, absolute rights (i.e., natural rights theory), the presence of human rationality and inviolable moral laws (i.e., Kantian categorical imperative), and the morality of good actors arising from unbiases behind a veil of ignorance (i.e., contractualism).

Ethics bluewashing: The implementation of superficial or misleading measures to appear ethical.

Ethics dumping: The export or import of unethical activities to a place with less strict regulations.

Ethics lobbying: The use of ethics to avoid or delay good and necessary regulation and enforcement.

Ethics shirking: The engaging of less ethical works over a period of time to lower the perceived resistance against such works.

Ethics shopping: The picking and choosing of ethics principles that are justified as a posteriori and retrofitted to pre-existing behaviors.

Explainability: This ethics principle relates to the lowering of opacity relating to data, AI algorithms and AI-driven decisions, the justification of its use, and the communication of details in a non-technical easy-to-understand manner to relevant stakeholders.

Fairness: This ethics principle relates to fair, equitable and appropriate educational practices that should be perpetuated by AI systems.

Human centricity: This ethics principle relates to the aim towards upholding human agency, dignity and autonomy, minimization of harm (and when necessary, weighed against a greater good), and equitable distribution of benefits.

Inclusivity: This ethics principle relates to inclusive and accessibility considerations applied to AI systems to meet different student needs in a personalized environment at scale.

Metaethics*:* The study of the nature (i.e., moral ontology), meaning (i.e., moral semantics), and the scope and knowledge to defend or support (i.e., moral epistemology) moral judgments.

Normative ethics: The study of the moral rules and standards that guide how individuals, institutions and societies should behave in a moral sense.

Privacy: This ethics principle relates to the protection of data subjects against injurious effects from the use of personal information applied in AI systems, without unduly affecting regulatory compliance tied to privacy and restricting AI development.

Trust: This ethics principle relates to the placing of confidence on AI systems and the provision of data to achieve educational objectives.

Virtue ethics: A type of normative ethics that emphasizes the inherent disposition of an individual, and not specific actions.