Proactive and reactive engagement of artificial intelligence methods for education: a review

Sruti Mallik*† and Ahana Gangopadhyay†

Washington University in St. Louis, St. Louis, MO, United States

The education sector has benefited enormously through integrating digital technology driven tools and platforms. In recent years, artificial intelligence based methods are being considered as the next generation of technology that can enhance the experience of education for students, teachers, and administrative staff alike. The concurrent boom of necessary infrastructure, digitized data and general social awareness has propelled these efforts further. In this review article, we investigate how artificial intelligence, machine learning, and deep learning methods are being utilized to support the education process. We do this through the lens of a novel categorization approach. We consider the involvement of AI-driven methods in the education process in its entirety—from students admissions, course scheduling, and content generation in the *proactive* planning phase to knowledge delivery, performance assessment, and outcome prediction in the *reactive* execution phase. We outline and analyze the major research directions under proactive and reactive engagement of AI in education using a representative group of 195 original research articles published in the past two decades, i.e., 2003–2022. We discuss the paradigm shifts in the solution approaches proposed, particularly with respect to the choice of data and algorithms used over this time. We further discuss how the COVID-19 pandemic influenced this field of active development and the existing infrastructural challenges and ethical concerns pertaining to global adoption of artificial intelligence for education.

KEYWORDS

artificial intelligence applications (AIA), artificial intelligence for education (AIEd), technology enhanced learning, machine learning, artificial intelligence for social good (AI4SG)

1. Introduction

Integrating computer-based technology and digital learning tools can enhance the learning experience for students and knowledge delivery process for educators (Lin et al., 2017; Mei et al., 2019). It can also help accelerate administrative tasks related to education (Ahmad et al., 2020). Therefore, researchers have continued to push the boundaries of including computer-based applications in classroom and virtual learning environments. Specifically in the past two decades, artificial intelligence (AI) based learning tools and technologies have received significant attention in this regard. In 2015, the United Nations General Assembly recognized the need to impart quality education at primary, secondary, technical, and vocational levels as one of their seventeen sustainable development goals or SDGs (United Nations, 2015). With this recognition, it is anticipated that research and development along the frontiers of including artificial intelligence for education will continue to be in the spotlight globally (Vincent-Lancrin and van der Vlies, 2020).

In the past there has been considerable discourse about how adoption of AI-driven methods for education might alter the course of how we perceive education (Dreyfus, 1999; Feenberg, 2017). However, in many of the earlier debates, the full potential of artificial intelligence was not recognized due to lack of supporting infrastructure. It was not until very recently that AI-powered techniques could be used in classroom environments. Since the beginning of the twenty-first century, there has been a rapid progress in the semiconductor industry in manufacturing chips that can handle computations at scale efficiently. In fact, in the coming decade too it is anticipated that this growth trajectory will continue with focus on wireless communication, data storage and computational resource development (Burkacky et al., 2022). With this parallel ongoing progress, using AI-driven platforms and tools to support students, educators, and policy-makers in education appears to be more feasible than ever.

The process of educating a student begins much before the student starts attending lectures and parsing lecture materials. In a traditional classroom education setup, administrative staff, and educators begin preparations related to making admissions decisions, scheduling of classes to optimize resources, curating course contents, and preliminary assignment materials several weeks prior to the term start date. In an online learning environment, similar levels of effort are put into structuring the course content and marketing the course availability to students. Once the term starts, the focus of educators is to deliver the course material, give out and grade assignments to assess progress and provide additional support to students who might benefit from that. The role of the students is to regularly acquire knowledge, ask clarifying questions and seek help to master the material. The role of administrative staff in this phase is less hands-on—they remain involved to ensure smooth and efficient overall progress. It is therefore a multi-step process involving many inter-dependencies and different stakeholders. Throughout this manuscript we refer to this multi-step process as the *end-to-end education process*.

In this review article, we review **how machine learning and artificial intelligence can be utilized in different phases of the end-to-end education process—from planning and scheduling to knowledge delivery and assessment. To systematically identify the different areas of active research with respect to engagement of AI in education, we first introduce a broad categorization of research articles in literature into those that address tasks prior to knowledge delivery and those that are relevant during** the process of knowledge delivery—i.e., *proactive vs. reactive engagement* with education. Proactive involvement of AI in education comes from its use in student admission logistics, curriculum design, scheduling and teaching content generation. Reactive involvement of AI is considerably broader in scope— AI-based methods can be used for designing intelligent tutoring systems, assessing performance and predicting student outcomes. In the schematic in Figure 1, we present an overview of our categorization approach. We have selected a sample set of research articles under each category and identified the key problem statements addressed using AI methods in the past 20 years. We believe that our categorization approach exposes to researchers the wide scope of using AI for the educational process. At the same time, it allows readers to identify the timeline of when certain

AI-driven tool might be applicable and what are the key challenges and concerns with using these tools at that time. The article further summarizes for expert researchers how the use of datasets and algorithms have evolved over the years and the scope for future research in this domain.

Through this review article, we aim to address the following questions:

- What were the widely studied applications of artificial intelligence in the end-to-end education process in the past two decades? How did the 2020 outbreak of the COVID-19 pandemic influence the landscape of research in this domain? Over the past two decades in retrospective view, has the usage of AI for education widened or bridged the gap between population groups with respect to access to quality education?
- How has the choice of datasets and algorithms in AI-driven tools and platforms evolved over this period—particularly in addressing the active research questions in the end-to-end education process?

The organization of this review article from here on is as follows. In Section 2, we define the scope of this review, outline the paper selection strategy and present the summary statistics. In Section 3, we contextualize our contribution in the light of technical review articles published in the domain of AIEd in the past 5 years. In Section 4, we present our categorization approach and review the scientific and technical contributions in each category. Finally, in Section 5, we discuss the major trends observed in research in the AIEd sector over the past two decades, discuss how the COVID-19 pandemic is reshaping the AIEd landscape and point out existing limitations in the global adoption of AI-driven tools for education. Additionally in Table 1, we provide a glossary of technical terms and their abbreviations that have been used throughout the paper.

2. Scope definition

The term artificial intelligence (AI) was coined in 1956 by John McCarthy (Haenlein and Kaplan, 2019). Since the first generally acknowledged work of McCulloch and Pitts in conceptualizing artificial neurons, AI has gone through several dormant periods and shifts in research focus. From algorithms that through exposure to somewhat noisy observational data learns to perform some pre-defined tasks, i.e., *machine learning (ML)* to more sophisticated approaches that learns the mapping of high-dimensional observations to representations in a lower dimensional space, i.e., *deep learning (DL)*—there is a plethora of computational techniques available currently. More recently, researchers and social scientists are increasingly using AI-based techniques to address social issues and to build toward a sustainable future (Shi et al., 2020). In this article, we focus on how one such social development aspect, i.e., education might benefit from usage of artificial intelligence, machine learning, and deep learning methods.

FIGURE 1
Overview of the categorization introduced in this review article.
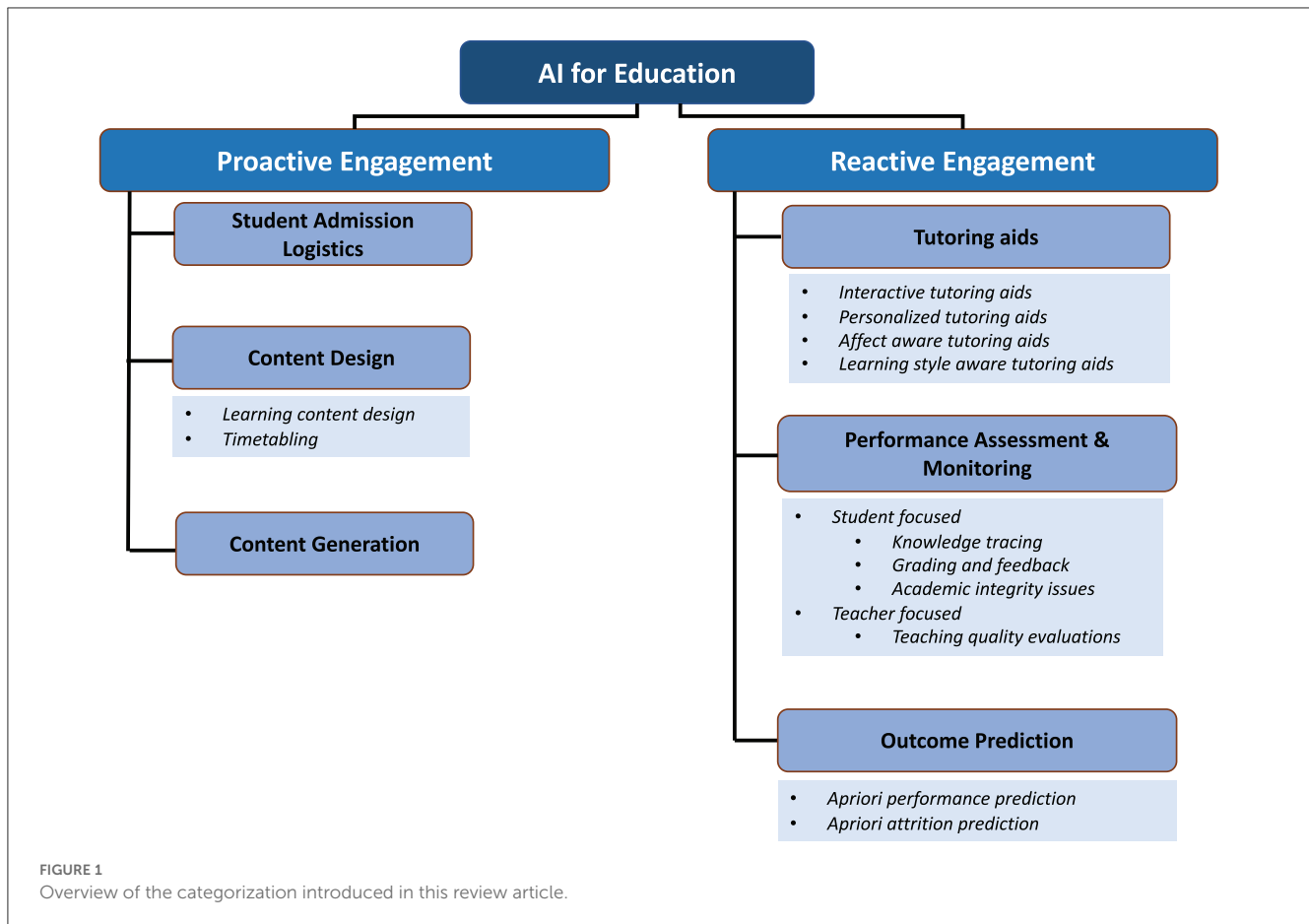
2.1. Paper search strategy

For the purpose of analyzing recent trends in this field (i.e., AIEd), we have sampled research articles published in peer-reviewed conferences and journals over the past 20 years, i.e. between 2003 and 2022, by leveraging the Google Scholar search engine. We identified our selected corpus of 195 research articles through a multi-step process. First, we identified a set of systematic review, survey papers and perspective papers published in the domain of artificial intelligence for education (AIEd) between the years of 2018 and 2022. To identify this list of review papers we used the keywords "artificial intelligence for education", "artificial intelligence for education review articles" and similar combinations in Google Scholar. We critically reviewed these papers and identified the research domains under AIEd that have received much attention in the past 20 years (i.e., 2002–2022) and that are closely tied to the end-to-end education process. Once, these research domains were identified, we further did a deep dive search using relevant keywords for each research area (for example, for the category tutoring aids, we used several keywords including intelligent tutoring systems, intelligent tutoring aids, computer-aided learning systems, affect-aware learning systems) to identify an initial set of technical papers in the sub-domain. We streamlined this initial set through the lens of significance of the problem statement, data used, algorithm proposed by thorough review of

each paper by both authors and retained the final set of 195 research articles.

2.2. Inclusion and exclusion criteria

Since the coinage of the term artificial intelligence, there is considerable debate in the scientific community about what is the scope of artificial intelligence. It is specifically challenging to delineate the boundaries as it is indeed a field that is subject to rapid technological change. Deep-dive analysis of this debate is beyond the scope of this paper. Instead, we have clearly stated in this section our inclusion/exclusion criteria with respect to selecting articles that surfaced in our search of involvement of AI for education. For this review article, we include research articles that use methods such as optimal search strategies (e.g., breadth-first search, depth-first search), density estimation, machine learning, Bayesian machine learning, deep learning and reinforcement learning. We do not include original research that proposes use of concepts and methods rooted in operations research, evolutionary algorithms, adaptive control theory, and robotics in our corpus of selected articles. In this review, we **only** consider peer-reviewed articles that were published in English. We do not include patented technologies and copyrighted EdTech software systems in our scope unless

peer-reviewed articles outlining the same contributions have been published by the authors.

2.3. Summary statistics

With the scope of our review defined above, here we provide the summary statistics of the 195 technical articles we covered in this review. In Figure 2, we show the distribution of the included scientific and technical articles over the past two decades. We also introspected the technical contributions in each category of

TABLE 1 Glossary of technical terms and their abbreviations frequently used in the paper.

Artificial Intelligence (AI): Simulation of human intelligence processes by machines.
Machine Learning (ML): Technologies or algorithms enabling computer systems to identify patterns from data, make decisions and improve their performance through experience.
Bayesian Machine Learning: A paradigm for constructing statistical models based on Bayes Theorem.
Deep Learning (DL): A class of machine learning algorithms that uses artificial neural networks consisting of multiple processing layers to map raw data into progressively higher-dimensional features.
Supervised Learning: A type of machine learning problem where algorithms are trained using labeled data points for the purpose of predicting labels for unseen examples.
Unsupervised Learning: A type of machine learning problem that learns patterns from unlabeled data.
Reinforcement Learning (RL): A type of machine learning problem where an agent learns an optimal set of actions in an environment through trial and feedback in order to maximize a reward.
Natural Language Processing (NLP): A branch of artificial intelligence and machine learning that enables computer systems to process and analyze natural language data in written or spoken format.
Convolutional Neural Networks (CNNs): A type of artificial neural network consisting of convolutional layers, most commonly used for processing visual imagery.
Generative AI: A type of artificial intelligence technology based on generative models that can produce text, images, audio and other kinds of content in response to prompts.

our categorization approach with respect to the target audiences they catered to (see Figure 3). We primarily identify target audience groups for educational technologies as such—pre-school students, elementary school students, middle and high school students, university students, standardized test examinees, students in e-learning platforms, students of MOOCs, and students in professional/vocational education. Articles where the audience group has not been clearly mentioned were marked as belonging to "Unknown" target audience category.

In Section 4, we introduce our categorization and perform a deep-dive to explore the breadth of technical contributions in each category. If applicable, we have further identified specific research problems currently receiving much attention as sub-categories within a category. In Table 2, we demonstrate the distribution of significant research problems within a category.

We defer the analysis of the identified trends from these summary plots to the Section 5 of this paper.

3. Related works

Artificial intelligence as a research area in technology has evolved gradually since 1950s. Similarly, the field of using computer based technology to support education has been actively developing since the 1980s. It is only however in the past few decades that there has been significant emphasis in adopting digital technologies including AI driven technologies in practice (Alam, 2021). Particularly, the introduction of open source generative AI algorithms, has spear-headed critical analyses of how AI can and should be used in the education sector (Baidoo-Anu and Owusu Ansah, 2023; Lund and Wang, 2023). In this backdrop of emerging developments, the number of review articles surveying the technical progress in the AIEd discipline has also increased in the last decade (see Figure 4). To generate Figure 4, we used Google Scholar as the search engine with the keywords artificial intelligence for education, artificial intelligence for education review articles and similar combinations using domain abbreviations. In this section, we discuss the premise of the review articles published in the *last 5*
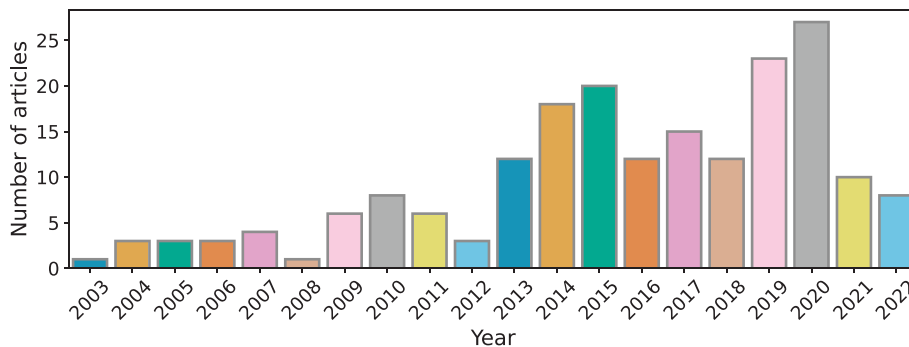
FIGURE 2
Distribution of the reviewed technical articles across the past two decades.
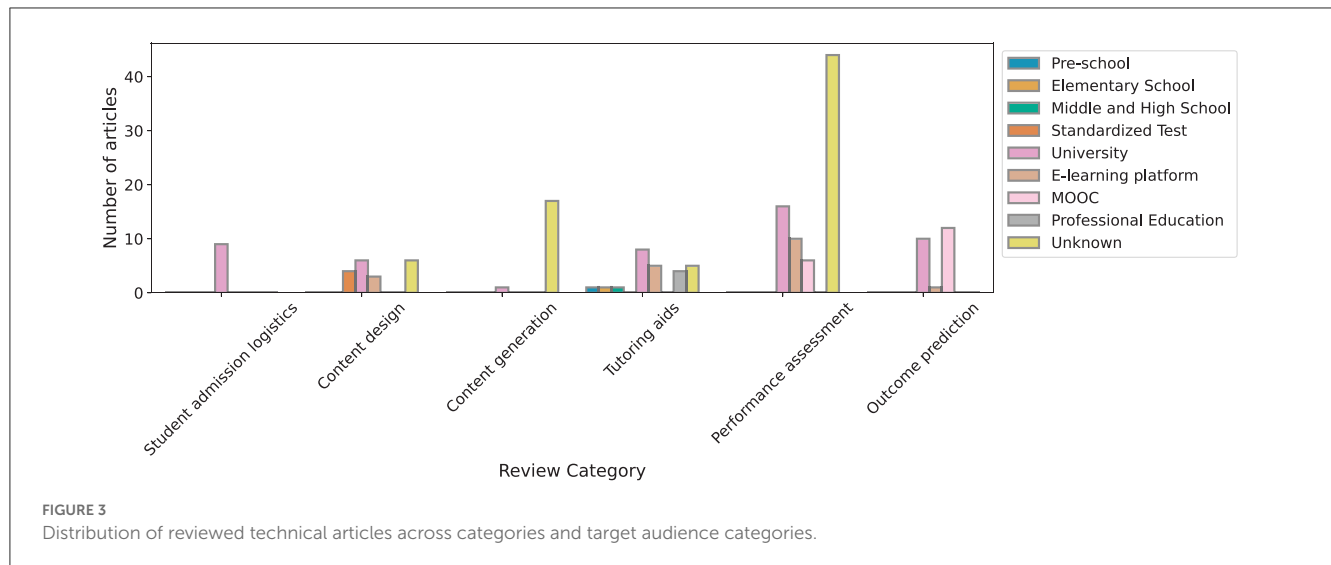
FIGURE 3
Distribution of reviewed technical articles across categories and target audience categories.

TABLE 2 Distribution of reviewed technical articles across sub-categories under each category.

Proactive vs. reactive engagement of AI	Review category	Review subcategory	Count
Proactive engagement	Student admission logistics	N/A	9
Proactive engagement	Content design	Learning content design	15
Proactive engagement	Content design	Timetabling	4
Proactive engagement	Content generation	N/A	22
Reactive engagement	Tutoring aids	Interactive tutoring aids	13
Reactive engagement	Tutoring aids	Personalized tutoring aids	8
Reactive engagement	Tutoring aids	Learning style based tutoring aids	7
Reactive engagement	Tutoring aids	Affect aware tutoring aids	5
Reactive engagement	Performance assessment	Student-focused	76
Reactive engagement	Performance assessment	Teacher-focused	9
Reactive engagement	Outcome prediction	Performance prediction	13
Reactive engagement	Outcome prediction	Drop-out prediction	14

years and situate this article with respect to previously published technical reviews.

Among the review articles identified based on the keyword search on Google Scholar and published between 2018 and 2022, one can identify two thematic categories—(i) *Technical reviews with categorization*: review articles that group research contributions based on some distinguishing factors, such as problem statement and solution methodology (Chassignol et al., 2018; Zawacki-Richter et al., 2019; Ahmad et al., 2020, 2022; Chen L. et al., 2020; Yufeia et al., 2020; Huang J. et al., 2021; Lameras and Arnab, 2021; Ouyang and Jiao, 2021; Zhai et al., 2021; Chen et al., 2022; Holmes and Tuomi, 2022; Namatherdhala et al., 2022; Wang and Cheng, 2022). (ii) *Perspectives on challenges, trends, and roadmap*: review articles that highlight the current state of research in a domain and offer critical analysis of the challenges and the future road map for the domain (Fahimirad and Kotamjani, 2018; Humble and Mozelius, 2019; Malik et al., 2019; Pedro et al.,

2019; Bryant et al., 2020; Hwang et al., 2020; Alam, 2021; Schiff, 2021). Closely linked with (i) are review articles that dive deep into the developments within a particular sub-category associated with AIEd, such as AIEd in the context of early childhood education (Su and Yang, 2022) and online higher education (Ouyang F. et al., 2022). We have designed this review article to belong to category (i). We distinguish between the different research problems in the context of AIEd through the lens of their timeline for engagement in the end-to-end education process and then perform a deeper review of ongoing research efforts in each category. To the best of our knowledge, such distinction between proactive and reactive involvement of AI in education along with an granular review of significant research questions in each category is presented for the first time through this paper (see schematic in Figure 1).

In Table 3, we have outlined the context of recently published technical reviews with categorization.
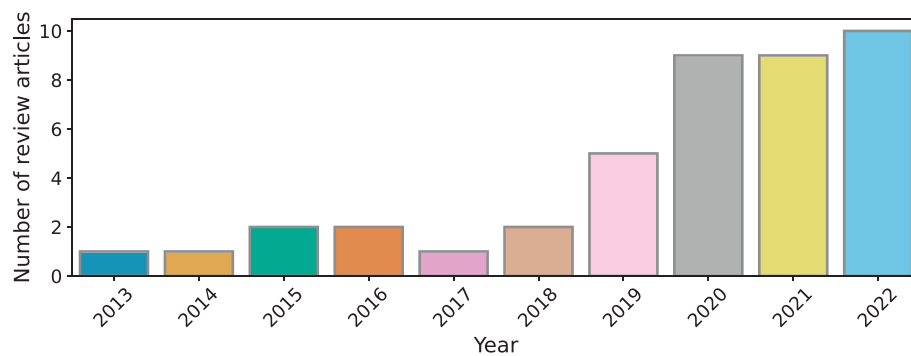
FIGURE 4
Number of review articles published in AIEd over the past decade.

4. Engaging artificial intelligence driven methods in stages of education

4.1. Proactive vs. reactive engagement of AI—An introduction

In the introductory section of this article, we have outlined how the process of education is a multi-step process and how it involves different stakeholders along the timeline. To this end, we can clearly identify that there are two distinct phases of engaging AI in the end-to-end education process. First, **proactive engagement** of AI—efforts in this phase are to design, curate and to ensure optimal use of resources, and second, **reactive engagement** of AI— efforts in this phase are to ensure that students acquire the necessary information and skills from the sessions they attend and provide feedback as needed.

In this review article, we distinguish between the scientific and technical contributions in the field of AIEd through the lens of these two distinct phases. This categorization is significant for the following reasons:

- First, through this hierarchical categorization approach, one can gauge the range of problems in the context of education that can be addressed using artificial intelligence. AI research related to personalized tutoring aids and systems has indeed had a head-start and is a mature area of research currently. However, the scope of using AI in the end-to-end education process is broad and rapidly evolving.
- Second, this categorization approach provides a retrospective overview of milestones achieved in AIEd through continuous improvement and enrichment of the data and algorithm leveraged in building AI models.
- Third, as this review touches upon both classroom and administrative aspect of education, readers can formulate a perspective for the myriad of infrastructural and ethical challenges that exist with respect to widespread adoption of AI-driven methods in education.

Within these broad categorizations, we further break down and analyze the research problems that have been addressed

using AI. For instance, in the proactive engagement phase, AI-based algorithms can be leveraged to determine student admission logistics, design curricula and schedules, and create course content. On the other hand, in the reactive engagement phase, AI-based methods can be used for designing intelligent tutoring systems (ITS), performance assessment, and prediction of student outcomes (see Figure 1). Another important distinction between the two phases lies in the nature of the available data to develop models. While the former primarily makes use of historical data points or pre-existing estimates of available resources and expectations about learning outcomes, the latter has at its disposal a growing pool of data points from the currently ongoing learning process, and can therefore be more adaptive and initiate faster pedagogical interventions to changing scopes and requirements.

4.2. Proactive engagement of AI for education

4.2.1. Student admission logistics

In the past, although a number of studies used statistical or machine learning-based approaches to analyze or model student admissions decisions, they had little role in the actual admissions process (Bruggink and Gambhir, 1996; Moore, 1998). However in the face of growing numbers of applicants, educational institutes are increasingly turning to AI-driven approaches to efficiently review applications and make admission decisions. For example, the Department of Computer Science at University of Texas Austin (UTCS) introduced an explainable AI system called GRADE (Graduate Admissions Evaluator) that uses logistic regression on past admission records to estimate the probability of a new applicant being admitted in their graduate program (Waters and Miikkulainen, 2014). While GRADE did not make the final admission decision, it reduced the number of full application reviews as well as review time per application by experts. Zhao et al. (2020) used features extracted from application materials of students as well as how they performed in the program of study to predict an incoming applicant's potential performance and identify students best suited for the program. An important metric for educational institutes with regard to student admissions

TABLE 3 Contextualization with respect to technical reviews published in the past 5 years (2018–2022).

Paper title	Summary
Artificial Intelligence trends in education: a narrative overview (Chassignol et al., 2018)	Categorizes AI in education into four categories—customized educational content, assessment and evaluation, adaptive systems and personalization, intelligent tutoring systems.
Systematic review of research on artificial intelligence applications in higher education—where are the educators (Zawacki-Richter et al., 2019)	Categorizes AI in education into four categories—profiling and prediction, assessment and evaluation, adaptive systems and personalization, intelligent tutoring systems.
Artificial Intelligence in Education: A Review (Chen L. et al., 2020)	Identifies and reviews four key ways in which AI has been adopted for education—automation of administrative processes and tasks, curriculum and content development, instruction, modeling students' learning process.
Review of the application of artificial intelligence in education (Yufeia et al., 2020)	Identifies and reviews aspects in which AI technology has been used in education—automatic grading system, interval reminder, teacher's feedback, virtual teachers, personalized learning, adaptive learning, augmented reality/virtual reality, accurate reading, intelligent campus and distance learning.
Artificial Intelligence in Education: A panoramic review (Ahmad et al., 2020)	Reviews the various applications of AI such as student grading and evaluations, students retention and drop out prediction, sentiment analysis, intelligent tutoring, classroom monitoring and recommendation systems.
A Review of Artificial Intelligence (AI) in Education from 2010 to 2020 (Zhai et al., 2021)	Reviews articles that use AI for social sciences such as in education and classifies the research questions into development layer (classification, matching, recommendation, and deep learning), application layer (feedback, reasoning, and adaptive learning), and integration layer (affection computing, role-playing, immersive learning, and gamification).
Artificial intelligence in education: the three paradigms (Ouyang and Jiao, 2021)	Identifies the paradigm shifts of AIEd and categorizes into AI-directed (learner-as-recipient), AI-supported (learner-as-collaborator), and AI-empowered (learner-as-leader).
Power to the teachers: an exploratory review on artificial intelligence in education (Lameras and Arnab, 2021)	Discusses research contribution along the five aspects of teaching and learning introduced by Dong and Chen (2020): 1. AIEd for preparing and transmitting learning content 2. AIEd for helping students to apply knowledge 3. AIEd for engaging students in learning tasks 4. AIEd for helping students to improvement through assessments and feedback 5. AIEd for helping students to become self-regulated learners.
A review on artificial intelligence in education (Huang J. et al., 2021)	Outlines the application of AI in education—adaptive learning, teaching evaluation, virtual classroom, smart campus, intelligent tutoring robots, and then analyzes its impact on teaching and learning.
Toward a tripartite research agenda: a scoping review of artificial intelligence in education research (Wang and Cheng, 2022)	Provides a scoping review of research studies on AIEd published between 2001 and 2021 and identifies and discusses three distinct agendas—Learning from AI, Learning about AI, and Learning with AI.
Two Decades of Artificial Intelligence in Education: contributors, Collaborations, Research Topics, Challenges, and Future Directions (Chen et al., 2022)	The authors identify the main research topics in AIEd in the past two decades to be—intelligent tutoring systems for special education, natural language processing for language education, educational robots for AI education, educational data mining for performance prediction, discourse analysis in computer-supported collaborative learning, neural networks for teaching evaluation, affective computing for learner emotion detection, and recommender systems for personalized learning.
Academic and Administrative Role of Artificial Intelligence in Education (Ahmad et al., 2022)	This review article aims to explore the academic and administrative applications of AI with an in-depth discussion on artificial intelligence applications in 1. Grading/Assessment 2. Admission 3. Virtual Reality (VR) for education 4. Learning Analytics.
A Comprehensive Overview of Artificial Intelligence Trends in Education (Namatherdhala et al., 2022)	The authors categorize application of AI for education into three distinct groups—Education administration, Instruction Design and Learning outcomes and briefly reviews each of them.
State of the art and practice in AI in education (Holmes and Tuomi, 2022)	The authors provide a review of existing AI systems in education and their pedagogic and educational assumptions. They also introduce a categorization for AIEd systems and discusses different ways of using AI in education and learning and different interpretations of what AI and education is or could be and existing roadblocks.

is yield rate, the rate at which accepted students decide to enroll at a given school. Machine learning has been used to predict enrollment decisions of students, which would help the institute make strategic admission decisions in order to improve their yield rate and optimize resource allocation (Jamison, 2017). Additionally, whether students enroll in suitable majors based on their specific backgrounds and prior academic performance is also indicative of future success. Machine learning has also been used to classify students into suitable majors in an attempt to set them up for academic success (Assiri et al., 2022).

Another research direction in this domain approaches the admissions problem from the perspective of students by predicting the probability that an applicant will get admission at a particular university in order to help applicants better target universities based on their profiles as well as university rankings (AlGhamdi et al., 2020; Goni et al., 2020; Mridha et al., 2022). Notably, more than one such work finds prior GPA (Grade Point Average) of students to be the most significant factor in admissions decisions (Young and Caballero, 2019; El Guabassi et al., 2021).

Given the high stakes involved and the significant consequences that admissions decisions have on the future of students, there has been considerable discourse on the ethical considerations of using AI in such applications, including its fairness, transparency, and privacy aspects (Agarwal, 2020; Finocchiaro et al., 2021). Aside from the obvious potential risks of worthy applicants getting rejected or unworthy applicants getting in, such systems

can perpetuate existing biases in the training data from human decision-making in the past (Bogina et al., 2022). For example, such systems might show unintentional bias toward certain demographics, gender, race, or income groups. Bogina et al. (2022) advocated for explainable models for making admission decisions, as well as proper system testing and balancing before reaching the end user. Emelianov et al. (2020) showed that demographic parity mechanisms like group-specific admission thresholds increase the utility of the selection process in such systems in addition to improving its fairness. Despite concerns regarding fairness and ethics, interestingly, university students in a recent survey rated algorithmic decision-making (ADM) higher than human decision-making (HDM) in admission decisions in both procedural and distributive fairness aspects (Marcinkowski et al., 2020).

4.2.2. Content design

In the context of education, we can define content as— (i) learning content for a course, curriculum, or test; and (ii) schedules/timetables of classes. We discuss AI/ML approaches for designing/structuring both of the above in this section.

(i) **Learning content design**: Prior to the start of the learning process, educators, and administrators are responsible for identifying an appropriate set of courses for a curriculum, an appropriate set of contents for a course, or an appropriate set of questions for a standardized test. In course and curriculum design, there is a large body of work using traditional systematic and relational approaches (Kessels, 1999), however the last decade saw several works using AI-informed curriculum design approaches. For example, Ball et al. (2019) uses classical ML algorithms to identify factors prior to declaration of majors in universities that adversely affect graduation rates, and advocates curriculum changes to alleviate these factors. Rawatlal (2017) uses tree-based approaches on historical records to prioritize the prerequisite structure of a curriculum in order to determine student progression routes that are effective. Somasundaram et al. (2020) proposes an Outcome Based Education (OBE) where expected outcomes from a degree program such as job roles/skills are identified first, and subsequently courses required to reach these outcomes are proposed by modeling the curriculum using ANNs. Doroudi (2019) suggests a semi-automated curriculum design approach by automatically curating low-cost, learner-generated content for future learners, but argues that more work is needed to explore data-driven approaches in curating pedagogically useful peer content.

For designing standardized tests such as TOEFL, SAT, or GRE, an essential criteria is to select questions having a consistent difficulty level across test papers for fair evaluation. This is also useful in classroom settings if teachers want to avoid plagiarism issues by setting multiple sets of test papers, or in designing a sequence of assignments or exams with increasing order of difficulty. This can be done through Question Difficulty Prediction (QDP) or Question Difficulty Estimation (QDE), an estimate of the skill level needed to answer a question correctly. QDP was historically estimated by pretesting on students or from expert ratings, which are expensive, time-consuming, subjective, and often vulnerable to leakage or exposure (Benedetto et al., 2022).

Rule-based algorithms relying on difficulty features extracted by experts were also proposed in Grivokostopoulou et al. (2014) and Perikos et al. (2016) for automatic difficulty estimation. As data-driven solutions became more popular, a common approach used linguistic features (Mothe and Tanguy, 2005; Stiller et al., 2016), readability scores, (Benedetto et al., 2020a; Yaneva et al., 2020), and/or word frequency features (Benedetto et al., 2020a,b; Yaneva et al., 2020) with ML algorithms such as linear regression, SVMs, tree-based approaches, and neural networks for downstream classification or regression, depending on the problem setup. With automatic testing systems and ready availability of large quantities of historical test logs, deep learning has been increasingly used for feature extraction (word embeddings, question representations, etc.) and/or difficulty estimation (Fang et al., 2019; Lin et al., 2019; Xue et al., 2020). Attention strategies have been used to model the difficulty contribution of each sentence in reading problems (Huang et al., 2017) or to model recall (how hard it is to recall the knowledge assessed by the question) and confusion (how hard it is to separate the correct answer from distractors) in Qiu et al. (2019). Domain adaptation techniques have also been proposed to alleviate the need of difficulty-labeled question data for each new course by aligning it with the difficulty distribution of a resource-rich course (Huang Y. et al., 2021). AlKhuzaey et al. (2021) points out that a majority of data-driven QDP approaches belong to language learning and medicine, possibly spurred on by the existence of a large number of international and national-level standardized language proficiency tests and medical licensing exams.

(ii) **Timetabling**: Educational Timetabling Problem (ETP) deals with the assignment of classes or exams to a limited number of time-slots such that certain constraints (e.g., availability of teachers, students, classrooms, and equipments) are satisfied. This can be divided into three types—course timetabling, school timetabling, and exam timetabling (Zhu et al., 2021). Timetabling not only ensures proper resource allocation, its design considerations (e.g., number of courses per semester, number of lectures per day, number of free time-slots per day) have noticeable impact on student attendance behavior and academic performance (Larabi-Marie-Sainte et al., 2021). Popular approaches in this domain such as mathematical optimization, meta-heuristic, hyper-heuristic, hybrid, and fuzzy logic approaches. Zhu et al. (2021) and Tan et al. (2021) mostly is beyond the scope of our paper (see Section 2.2). Having said that, it must be noted that machine learning has often been used in conjunction with such mathematical techniques to obtain better performing algorithms. For example, Kenekayoro (2019) used supervised learning to find approximations for evaluating solutions to optimization problems—a critical step in heuristic approaches. Reinforcement learning has been used to select low-level heuristics in hyper-heuristic approaches (Obit et al., 2011; Özcan et al., 2012) or to obtain a suitable search neighborhood in mathematical optimization problems (Goh et al., 2019).

4.2.3. Content generation

The difference between content design and content generation is that of curation versus creation. While the former focuses on selecting and structuring the contents for a course/curriculum in a

way most appropriate for achieving the desired learning outcomes, the latter deals with generating the course material itself. AI has been widely adopted to generate and improve learning content prior to the start of the learning process, as discussed in this section.

Automatically generating questions from narrative or informational text, or automatically generating problems for analytical concepts are becoming increasingly important in the context of education. Automatic question generation (AQG) from teaching material can be used to improve learning and comprehension of students, assess information retention from the material and aid teachers in adding Supplementary material from external sources without the time-intensive process of authoring assessments from them. They can also be used as a component in intelligent tutoring systems to drive engagement and assess learning. AQG essentially consists of two aspects: content selection or *what to ask*, and question construction or *how to ask it* (Pan et al., 2019), traditionally considered as separate problems. Content selection for questions was typically done using different statistical features (sentence length, word/sentence position, word frequency, noun/pronoun count, presence of superlatives, etc.) (Agarwal and Mannem, 2011) or NLP techniques such as syntactic or semantic parsing (Heilman, 2011; Lindberg et al., 2013), named entity recognition (Kalady et al., 2010) and topic modeling (Majumder and Saha, 2015). Machine learning has also been used in such contexts, e.g., to classify whether a certain sentence is suitable to be used as a stem in cloze questions (passage with a portion occluded which needs to be replaced by the participant) (Correia et al., 2012). The actual question construction, on the other hand, traditionally adopted rule-based methods like transformation-based approaches (Varga and Ha, 2010) or template-based approaches (Mostow and Chen, 2009). The former rephrased the selected content using the correct question key-word after deleting the target concept, while the latter used pre-defined templates that can each capture a class of questions. Heilman and Smith (2010) used an overgenerate-and-rank approach to overgenerate questions followed by the use of supervised learning for ranking them, but still relied on handcrafted generating rules. Following the success of neural language models and concurrent with the release of large-scale machine reading comprehension datasets (Nguyen et al., 2016; Rajpurkar et al., 2016), question generation was later framed as a sequence-to-sequence learning problem that directly maps a sentence (or the entire passage containing the sentence) to a question (Du et al., 2017; Zhao et al., 2018; Kim et al., 2019), and can thus be trained in an end-to-end manner (Pan et al., 2019). Reinforcement learning based approaches that exploit the rich structural information in the text have also been explored in this context (Chen Y. et al., 2020). While text is the most common type of input in AQG, such systems have also been developed for structured databases (Jouault and Seta, 2013; Indurthi et al., 2017), images (Mostafazadeh et al., 2016), and videos (Huang et al., 2014), and are typically evaluated by experts on the quality of generated questions in terms of relevance, grammatical, and semantic correctness, usefulness, clarity etc.

Automatically generating problems that are similar to a given problem in terms of difficulty level, can greatly benefit teachers in setting individualized practice problems to avoid plagiarism and still ensure fair evaluation (Ahmed et al., 2013). It also enables the students to be exposed to as many (and diverse) training exercises as needed in order to master the underlying concepts (Keller, 2021). In this context, mathematical word problems (MWPs)—an established way of inculcating math modeling skills in K-12 education—have witnessed significant research interest. Preliminary work in automatic MWP generation take a template-based approach, where an existing problem is generalized into a template, and a solution space fitting this template is explored to generate new problems (Deane and Sheehan, 2003; Polozov et al., 2015; Koncel-Kedziorski et al., 2016). Following the same shift as in AQG, Zhou and Huang (2019) proposed an approach using Recurrent Neural Networks (RNNs) that encodes math expressions and topic words to automatically generate such problems. Subsequent research along this direction has focused on improving topic relevance, expression relevance, language coherence, as well as completeness and validity of the generated problems using a spectrum of approaches (Liu et al., 2021; Wang et al., 2021; Wu et al., 2022).

On the other end of the content generation spectrum lie systems that can generate solutions based on the content and related questions, which include Automatic Question Answering (AQA) systems, Machine Reading Comprehension (MRC) systems and automatic quantitative reasoning problem solvers (Zhang D. et al., 2019). These have achieved impressive breakthroughs with the research into large language models and are widely regarded in the larger narrative as a stepping-stone toward Artificial General Intelligence (AGI), since they require sophisticated natural language understanding and logical inferencing capabilities. However, their applicability and usefulness in educational settings remains to be seen.

4.3. Reactive engagement of AI for education

4.3.1. Tutoring aids

Technology has been used to aid learners to achieve their learning goals for a long time. More focused effort on developing computer-based tutoring systems in particular started following the findings of Bloom (Bloom, 1984)—students who received tutoring in addition to group classes fared two standard deviations better than those who only participated in group classes. Given its early start, research on Intelligent Tutoring Systems (ITS) is relatively more mature than other research areas under the umbrella of AIEd research. Fundamentally, the difference between designs of ITS comes from the difference in the *underlying assumption of what augments the knowledge acquisition process for a student*. In the review paper on ITS (Alkhatlan and Kalita, 2018), a comprehensive timeline and overview of research in this domain is provided. Instead of repeating findings from previous reviews under this category, we distinguish between ITS designs through the lens of the underlying hypotheses. We primarily identified four hypotheses that are currently receiving much attention from the research community—emphasis on tutor-tutee interaction, emphasis of personalization, inclusion of affect and emotion, and consideration of specific learning styles. It must be noted that tutoring itself is

an interactive process, therefore most designs in this category have a basic interactive setup. However, contributions in categories (ii) through (iv), have other concept as the focal point of their tutoring aid design.

(i) **Interactive tutoring aids**: Previous research in education (Jackson and McNamara, 2013) has pointed out that *when a student is actively interacting with the educator or the course contents, the student stays engaged in the learning process for a longer duration.* Learning systems that leverage this hypothesis can be categorized as interactive tutoring aids. These frameworks allow the student to communicate (verbally or through actions) with the teacher or the teaching entity (robots or software) and get feedback or instructions as needed.

Early designs of interactive tutoring aids for teaching and support comprised of rule-based systems mirroring interactions between expert teacher and student (Arroyo et al., 2004; Olney et al., 2012) or between peer companions (Movellan et al., 2009). These template rules provided output based on the inputs from the student. Over the course of time, interactive tutoring systems gradually shifted to inferring the student's state in real time from the student's interactions with the tutoring system and providing fine-tuned feedback/instructions based on the inference. For instance, Gordon and Breazeal (2015) used a Bayesian active learning algorithm to assess student's word reading skills while the student was being taught by a robot. Presently, a significant number of frameworks belonging to this category uses chatbots as a proxy for a teacher or a teaching assistant (Ashfaque et al., 2020). These recent designs can use a wide variety of data such as text and speech, and rely on a combination of sophisticated and resource-intensive deep-learning algorithms to infer and further customize interactions with the student. For example, Pereira (2016) presents "@dawebot" that uses NLP techniques to train students using multiple choice question quizzes. Afzal et al. (2020) presents a conversational medical school tutor that uses NLP and natural language understanding (NLU) to understand user's intent and present concepts associated with a clinical case.

Hint construction and partial solution generation is yet another method to keep students engaged interactively. For instance, Green et al. (2011) used Dynamic Bayes Nets to construct a curriculum of hints and associated problems. Wang and Su (2015) in their architecture iGeoTutor assisted students in mastering geometry theorems by implementing search strategies (e.g., DFS) from partially complete proofs. Pande et al. (2021) aims to improve individual and self-regulated learning in group assignments through a conversational system built using NLU and dialogue management systems that prompts the students to reflect on lessons learnt while directing them to partial solutions.

One of the requirements of certain professional and vocational training such as biology, medicine, military etc. is practical experience. With the support of booming infrastructure, many such training programs are now adopting AI-driven augmented reality (AR)/virtual reality (VR) lesson plans. Interconnected modules driven by computer vision, NLU, NLP, text-to-speech (TTS), information retrieval algorithms facilitate lessons and/or assessments in biology (Ahn et al., 2018), surgery and medicine (Mirchi et al., 2020), pathological laboratory analysis (Taoum et al., 2016), and military leadership training (Gordon et al., 2004).

(ii) **Personalized tutoring aids**: As every student is unique, *personalizing instruction and teaching content can positively impact the learning outcome of the student* (Walkington, 2013)— tutoring systems that incorporate this can be categorized as personalized learning systems or personalized tutoring aids. Notably, personalization during instruction can occur through course content sequencing and display of prompts and additional resources among others.

The sequence in which a student reviews course topics plays an important role in their mastery of a concept. One of the criticisms of early computer based learning tools was the "one approach fits all" method of execution. To improve upon this limitation, personalized instructional sequencing approaches were adopted. In some early developments, Idris et al. (2009) developed a course sequencing method that mirrored the role of an instructor using soft computing techniques such as self organized maps and feed-forward neural networks. Lin et al. (2013) propose the use of decision trees trained on student background information to propose personalized learning paths for creativity learning. Reinforcement learning (RL) naturally lends itself to this task. Here an optimal policy (sequence of instructional activities) is inferred depending on the cognitive state of a student (estimated through knowledge tracing) in order to maximize a learning-related reward function. As knowledge delivery platforms are increasingly becoming virtual and thereby generating more data, deep reinforcement learning has been widely applied to the problem of instructional sequencing (Reddy et al., 2017; Upadhyay et al., 2018; Pu et al., 2020; Islam et al., 2021). Doroudi (2019) presents a systematic review of RL-induced instructional policies that were evaluated on students, and concludes that over half outperform all baselines they were tested against.

In order to display a set of relevant resources personalized with respect to a student state, algorithmic search is carried out in a knowledge repository. For instance, Kim and Shaw (2009) uses information retrieval and NLP techniques to present two frameworks: PedaBot that allows students to connect past discussions to the current discussion thread and MentorMatch that facilitates student collaboration customized based on student's current needs. Both PedaBot and MentorMatch systems use text data coming from a live discussion board in addition to textbook glossaries. In order to reduce information overload and allow learners to easily navigate e-learning platforms, Deep Learning-Based Course Recommender System (DECOR) has been proposed recently (Li and Kim, 2021)—this architecture comprises of neural network based recommendation systems trained using student behavior and course related data.

(iii) **Affect aware tutoring aids**: Scientific research proposes *incorporating affect and behavioral state of the learner into the design of the tutoring system as it enhances the effectiveness of the teaching process* (Woolf et al., 2009; San Pedro et al., 2013). Arroyo et al. (2014) suggests that cognition, meta-cognition and affect should indeed be modeled using real time data and used to design intervention strategies. Affect and behavioral state of a student can generally be inferred from sensor data that tracks minute physical movements of the student (eyegaze, facial expression, posture etc.). While initial approaches in this direction required sensor data, a major constraint for availing and using such data pertains to

ethical and legal reasons. "Sensor-free" approaches have thereby been proposed that use data such as student self-evaluations and/or interaction logs of the student with the tutoring system. Arroyo et al. (2010) and Woolf et al. (2010) use interaction data to build affect detector models—the raw data in these cases are first distilled into meaningful features and then fed into simple classifier models that detect individual affective states. DeFalco et al. (2018) compares the usage of sensor and interaction data in delivering motivational prompts in the course of military training. In Botelho et al. (2017), uses RNNs to enhance the performance of sensor-free affect detection models. In their review of affect and emotion aware tutoring aids, Harley et al. (2017) explore in depth the different use cases for affect aware intelligent tutoring aids such as enriching user experience, better curating learning material and assessments, delivering prompts for appraisal, navigational instructions etc., and the progress of research in each direction.

(iv) **Learning style aware tutoring aids**: Yet another perspective in the domain of ITS *pertains to customizing course content according to learning styles of students for better end outcomes*. Kolb (1976), Pask (1976), Honey and Mumford (1986), and Felder (1988) among others proposed different approaches to categorize learning styles of students. Traditionally, an individual's learning style was inferred via use of a self-administered questionnaire. However, more recently machine learning based methods are being used to categorize learning styles more efficiently from noisy subject data. Lo and Shu (2005), Villaverde et al. (2006), Alfaro et al. (2018), and Bajaj and Sharma (2018) use as input the completed questionnaire and/or other data sources such as interaction data and behavioral data of students, and feed the extracted features into feed-forward neural networks for classification. Unsupervised methods such as self-organizing map (SOM) trained using curated features have also been used for automatic learning style identification (Zatarain-Cabada et al., 2010). While for categorization per the Felder and Silverman learning style model, count of student visits to different sections of the e-learning platform are found to be more informative (Bernard et al., 2015; Bajaj and Sharma, 2018), for categorization per the Kolb learning model, student performance, and student preference features were found to be more relevant. Additionally, machine learning approaches have also been proposed for learning style based learning path design. In Mota (2008), learning styles are first identified through a questionnaire and represented on a polar map, thereafter neural networks are used to predict the best presentation layout of the learning objective for a student. It is worthwhile to point out, however, that in recent years instead of focusing on customizing course content with respect to certain pre-defined learning styles, more research efforts are focused on curating course material based on how an individual's overall preferences vary over time (Chen and Wang, 2021).

4.3.2. Performance assessment and monitoring

A critical component of the knowledge delivery phase involves assessing student performance by tracing their knowledge development and providing grades and/or constructive feedback on assignments and exams, while simultaneously ensuring academic integrity is upheld. Conversely, it is also important to evaluate the quality and effectiveness of teaching, which has a tangible impact on the learning outcomes of students. AI-driven performance assessment and monitoring tools have been widely developed for both learners and educators. Since a majority of evaluation material are in textual format, NLP-based models in particular have a major presence in this domain. We divide this section into student-focused and teacher-focused approaches, depending on the direct focus group of such applications.

(i) **Student-focused**:

Knowledge tracing. An effective way of monitoring the learning progress of students is through knowledge tracing, which models knowledge development in students in order to predict their ability to answer the next problem correctly given their current mastery level of knowledge concepts. This not only benefits the students by identifying areas they need to work on, but also the educators in designing targeted exercises, personalized learning recommendations and adaptive teaching strategies (Liu et al., 2019). An important step of such systems is cognitive modeling, which models the latent characteristics of students based on their current knowledge state. Traditional approaches for cognitive modeling include factor analysis methods which estimate student knowledge by learning a function (logistic in most cases) based on various factors related to the students, course materials, learning and forgetting behavior, etc. (Pavlik and Anderson, 2005; Cen et al., 2006; Pavlik et al., 2009). Another research direction explores Bayesian inference approaches that update student knowledge states using probabilistic graphical models like Hidden Markov Model (HMM) on past performance records (Corbett and Anderson, 1994), with substantial research being devoted to personalizing such model parameters based on student ability and exercise difficulty (Yudelson et al., 2013; Khajah et al., 2014). Recommender system techniques based on matrix factorization have also been proposed, which predict future scores given a student-exercise performance matrix with known scores (Thai-Nghe et al., 2010; Toscher and Jahrer, 2010). Abdelrahman et al. (2022) provides a comprehensive taxonomy of recent work in deep learning approaches for knowledge tracing. Deep knowledge tracing (DKT) was one of the first such models which used recurrent neural network architectures for modeling the latent knowledge state along with its temporal dynamics to predict future performance (Piech et al., 2015a). Extensions along this direction include incorporating external memory structures to enhance representational power of knowledge states (Zhang et al., 2017; Abdelrahman and Wang, 2019), incorporating attention mechanisms to learn relative importance of past questions in predicting current response (Pandey and Karypis, 2019; Ghosh et al., 2020), leveraging textual information from exercise materials to enhance prediction performance (Su et al., 2018; Liu et al., 2019) and incorporating forgetting behavior by considering factors related to timing and frequency of past practice opportunities (Nagatani et al., 2019; Shen et al., 2021). Graph neural network based architectures were recently proposed in order to better capture dependencies between knowledge concepts or between questions and their underlying knowledge concepts (Nakagawa et al., 2019; Tong et al., 2020; Yang et al., 2020). Specific to programming, Wang et al. (2017) used a sequence of embedded program submissions to train RNNs to predict performance in the

current or the next programming exercise. However as pointed out in Abdelrahman et al. (2022), handling of non-textual content as in images, mathematical equations or code snippets to learn richer embedding representations of questions or knowledge concepts remains relatively unexplored in the domain of knowledge tracing.

Grading and feedback. While technological developments have made it easier to provide content to learners at scale, scoring their submitted work and providing feedback on similar scales remains a difficult problem. While assessing multiple-choice and fill-in-the-blank type questions is easy enough to automate, automating assessment of open-ended questions (e.g., short answers, essays, reports, code samples) and questions requiring multi-step reasoning (e.g., theorem proving, mathematical derivations) is equally hard. But automatic evaluation remains an important problem not only because it reduces the burden on teaching assistants and graders, but also removes grader-to-grader variability in assessment and helps accelerate the learning process for students by providing real-time feedback (Srikant and Aggarwal, 2014).

In the context of written prose, a number of Automatic Essay Scoring (AES) and Automatic Short Answer Grading (ASAG) systems have been developed to reliably evaluate compositions produced by learners in response to a given prompt, and are typically trained on a large set of written samples pre-scored by expert raters (Shermis and Burstein, 2003; Dikli, 2006). Over the last decade, AI-based essay grading tools evolved from using handcrafted features such as word/sentence count, mean word/sentence length, n-grams, word error rates, POS tags, grammar, and punctuation (Adamson et al., 2014; Phandi et al., 2015; Cummins et al., 2016; Contreras et al., 2018) to automatically extracted features using deep neural network variants (Taghipour and Ng, 2016; Dasgupta et al., 2018; Nadeem et al., 2019; Uto and Okano, 2020). Such systems have been developed not only to provide holistic scoring (assessing essay quality with a single score), but also for more fine-grained evaluation by providing scoring along specific dimensions of essay quality, such as organization (Persing et al., 2010), prompt-adherence (Persing and Ng, 2014), thesis clarity (Persing and Ng, 2013), argument strength (Persing and Ng, 2015), and thesis strength (Ke et al., 2019). Since it is often expensive to obtain expert-rated essays to train on each time a new prompt is introduced, considerable attention has been given to cross-prompt scoring using multi-task, domain adaptation, or transfer learning techniques, both with handcrafted (Phandi et al., 2015; Cummins et al., 2016) and automatically extracted features (Li et al., 2020; Song et al., 2020). Moreover, feedback being a critical aspect of essay drafting and revising, AES systems are increasingly being adopted into Automated Writing Evaluation (AWE) systems that provide formative feedback along with (or instead of) final scores and therefore have greater pedagogical usefulness (Hockly, 2019). For example, AWE systems have been developed for providing feedback on errors in grammar, usage and mechanics (Burstein et al., 2004) and text evidence usage in response-to-text student writings (Zhang H. et al., 2019).

AI-based evaluation tools are also heavily used in computer science education, particularly programming, due to its inherent structure and logic. Traditional approaches for automated grading of source codes such as test-case based assessments (Douce et al.,

2005) and assessments using code metrics (e.g., lines of code, number of variables, number of statements), while simple, are neither robust nor effective at evaluating program quality.

A more useful direction measures similarities between abstract representations (control flow graphs, system dependence graphs) of the student's program and correct implementations of the program (Wang et al., 2007; Vujošević-Janičić et al., 2013) for automatic grading. Such similarity measurements could also be used to construct meaningful clusters of source codes and propagate feedback on student submissions based on the cluster they belong to (Huang et al. (2013); Mokbel et al. (2013). Srikant and Aggarwal (2014) extracts informative features from abstract representations of the code to train machine learning models using expert-rated evaluations in order to output a finer-grained evaluation of code quality. Piech et al. (2015b) used RNNs to learn program embeddings that can be used to propagate human comments on student programs to orders of magnitude more submissions. A bottleneck in automatic program evaluation is the availability of labeled code samples. Approaches proposed to overcome this issue include learning question-independent features from code samples (Singh et al., 2016; Tarcsay et al., 2022) or zero-shot learning using human-in-the-loop rubric sampling (Wu et al., 2019).

Elsewhere, driven by the maturing of automatic speech recognition technology, AI-based assessment tools have been used for mispronunciation detection in computer-assisted language learning (Li et al., 2009, 2016; Zhang et al., 2020) or the more complex problem of spontaneous speech evaluation where the student's response is not known *apriori* (Shashidhar et al., 2015). Mathematical language processing (MLP) has been used for automatic assessment of open response mathematical questions (Lan et al., 2015; Baral et al., 2021), mathematical derivations (Tan et al., 2017), and geometric theorem proving (Mendis et al., 2017), where grades for previously unseen student solutions are predicted (or propagated from expert-provided grades), sometimes along with partial credit assignment. Zhang et al. (2022), moreover, overcomes the limitation of having to train a separate model per question by using multi-task and meta-learning tools that promote generalizability to previously unseen questions.

Academic integrity issues. Another aspect of performance assessment and monitoring is to ensure the upholding of academic integrity by detecting plagiarism and other forms of academic or research misconduct. Foltýnek et al. (2019) in their review paper on academic plagiarism detection in text (e.g., essays, reports, research papers) classifies plagiarism forms according to an increasing order of obfuscation level, from verbatim and near-verbatim copying to translation, paraphrasing, idea-preserving plagiarism, and ghostwriting. In a similar fashion, plagiarism detection methods have been developed for increasingly complex types of plagiarism, and widely adopt NLP and ML-based techniques for each (Foltýnek et al., 2019). For example, lexical detection methods use n-grams (Alzahrani, 2015) or vector space models (Vani and Gupta, 2014) to create document representations that are subsequently thresholded or clustered (Vani and Gupta, 2014) to identify suspicious documents. Syntax-based methods rely on Part-of-speech (PoS) tagging (Gupta et al., 2014), frequency of PoS tags (Hürlimann et al., 2015), or comparison of syntactic trees (Tschuggnall and Specht, 2013). Semantics-based methods

employ techniques such as word embeddings (Ferrero et al., 2017), Latent Semantic Analysis (Soleman and Purwarianti, 2014), Explicit Semantic Analysis (Meuschke et al., 2017), and word alignment (Sultan et al., 2014), often in conjunction with other ML-based techniques for downstream classification (Alfikri and Purwarianti, 2014; Hänig et al., 2015). Complementary to such textual analysis-based methods, approaches that use non-textual elements like citations, math expressions, figures, etc. also adopt machine learning for plagiarism detection (Pertile et al., 2016). Foltỳnek et al. (2019) also provides a comprehensive summary of how classical ML algorithms such as tree-based methods, SVMs and neural networks have been successfully used to combine more than one type of detection method to create the best-performing meta-system. More recently, deep learning models such as different variants of convolutional and recurrent neural network architectures have also been used for plagiarism detection (El Mostafa Hambi, 2020; El-Rashidy et al., 2022).

In computer science education where programming assignments are given to evaluate students, source code plagiarism can also been classified based on increasing levels of obfuscation (Faidhi and Robinson, 1987). The detection process typically involves transforming the code into a high-dimensional feature representation followed by measurement of code similarity. Aside from traditionally used features extracted based on structural or syntactic properties of programs (Ji et al., 2007; Lange and Mancoridis, 2007), NLP-based approaches such as n-grams (Ohmann and Rahal, 2015), topic modeling (Ullah et al., 2021), character and word embeddings (Manahi, 2021), and character-level language models (Katta, 2018) are increasingly being used for robust code representations. Similarly for downstream similarity modeling or classification, unsupervised (Acampora and Cosma, 2015) and supervised (Bandara and Wijayarathna, 2011; Manahi, 2021) machine learning and deep learning algorithms are popularly used.

It is worth noting that AI itself makes plagiarism detection an uphill battle. With the increasing prevalence of easily accessible large language models like InstructGPT (Ouyang L. et al., 2022) and ChatGPT (Blog, 2022) that are capable of producing natural-sounding essays and short answers, and even working code snippets in response to a text prompt, it is now easier than ever for dishonest learners to misuse such systems for authoring assignments, projects, research papers or online exams. How plagiarism detection approaches, along with teaching and evaluation strategies, evolve around such systems remains to be seen.

(ii) **Teacher-focused**: Teaching Quality Evaluations (TQEs) are important sources of information in determining teaching effectiveness and in ensuring learning objectives are being met. The findings can be used to improve teaching skills through appropriate training and support, and also play a significant role in employment and tenure decisions and the professional growth of teachers. Such evaluations have been traditionally performed by analyzing student evaluations, teacher mutual evaluations, teacher self-evaluations and expert evaluations (Hu, 2021), which are labor-intensive to analyze at scale. Machine learning and deep learning algorithms can help with teacher evaluation by performing sentiment analysis

of student comments on teacher performance (Esparza et al., 2017; Gutiérrez et al., 2018; Onan, 2020), which provides a snapshot of student attitudes toward teachers and their overall learning experiences. Further, such quantified sentiments and emotional valence scores have been used to predict students' recommendation scores for teachers in order to determine prominent factors that influence student evaluations (Okoye et al., 2022). Vijayalakshmi et al. (2020) uses student ratings related to class planning, presentation, management, and student participation to directly predict instructor performance.

Apart from helping extract insights from teacher evaluations, AI can also be used to evaluate teaching strategies on the basis of other data points from the learning process. For example, Duzhin and Gustafsson (2018) used a symbolic regression-based approach to evaluate the impact of assignment structures and collaboration type on student scores, which course instructors can use for the purpose of self-evaluation. Several works use a combination of student ratings and attributes related to the course and the instructor to predict instructor performance and investigate factors affecting learning outcomes (Mardikyan and Badur, 2011; Ahmed et al., 2016; Abunasser et al., 2022) .

4.3.3. Outcome prediction

While a course is ongoing, one way to assess knowledge development in students is through graded assignments and projects. On the other hand, educators can also benefit from automatic prediction of students' performance and automatic identification of students at risk of course non-completion. This can be accomplished by monitoring students' patterns of engagement with the course material in association with their demographic information. Such *apriori* understanding of a student's outcome allows for designing effective intervention strategies. Presently, most K-12, undergraduate and graduate students, when necessary resources are available, rely on computer and web-based infrastructure (Bulman and Fairlie, 2016). A rich source of data indicating student state is therefore generated when a student interacts with the course modules. Prior to computers being such an integral component in education, researchers frequently used surveys and questionnaires to gauge student engagement, sentiment, and attrition probability. In this section we will summarize research developments in the field of AI that *generate early prediction of student outcomes—both final performance and possibility of drop-out.*

Early research in outcome prediction focused on building explanatory regression-based models for understanding student retention using college records (Dey and Astin, 1993). The active research direction in this space gradually shifted to tackling the more complex and more actionable problems of understanding whether a student will complete a program (Dekker et al., 2009), estimating the time a student will take to complete a degree (Herzog, 2006) and predicting the final performance of a student (Nghe et al., 2007) given the current student state. In the subsequent paragraphs, we will be discussing the research contributions for outcome prediction with distinction between performance prediction in assessments and course attrition prediction. Note that

we discuss these separately as poor performance in any assessment cannot be generalized into a course non-completion.

(i) *Apriori* **performance prediction**: *Apriori* prediction of performance of a student has several benefits—it allows a student to evaluate their course selection, and allows educators to evaluate progress and offer additional assistance as needed. Not surprisingly therefore AI-based methods have been proposed to automate this important task in the education process.

Initial research articles predicting performance estimated time to degree completion (Herzog, 2006) using student demographic, academic, residential and financial aid information, student parent data and school transfer records. In a related theme, researchers have also mapped the question of performance prediction into a final exam grade prediction problem (e.g., excellent, good, fair, fail; Nghe et al., 2007; Bydžovská, 2016; Dien et al., 2020). This granular prediction eventually allows educators to assess which students require additional tutoring. Baseline algorithms in this context are Decision Trees, Support Vector Machines, Random Forests, Artificial Neural Networks etc. (regression or classification based on the problem setup). Researchers have aimed to improve the performance of the predictors by including relevant information such as student engagement, interactions (Ramesh et al., 2013; Bydžovská, 2016), role of external incentives (Jiang et al., 2014), and previous performance records (Tamhane et al., 2014). Xu et al. (2017) proposed that a student's performance or when the student anticipates graduation should be predicted progressively (using an ensemble machine learning method) over the duration of the student's tenure as the academic state of the student is ever-evolving and can be traced through their student records. The process of generalizing performance prediction to non-traditional modes of learning such as hybrid or blended learning and on-line learning has benefitted from the inclusion of additional information sources such as web-browsing information (Trakunphutthirak et al., 2019), discussion forum activity and student study habits (Gitinabard et al., 2019).

In addition to exploring a more informative and robust feature set, recently, deep learning based approaches have been identified to outperform traditional machine learning algorithms. For example, Waheed et al. (2020) used deep feed-forward neural networks and split the problem of predicting student grade into multiple binary classification problems viz., Pass-Fail, Distinction-Pass, Distinction-Fail, Withdrawn-Pass. Tsiakmaki et al. (2020) analyzed if transfer learning (i.e., pre-training neural networks on student data on a different course) can be used to accurately predict student performance. Chui et al. (2020) used a generative adversarial network based architecture, to address the challenges of low volume of training data in alternative learning paradigms such as supportive learning. Dien et al. (2020) proposed extensive data pre-processing using min-max scaler, quantile transformation, etc. before passing the data in a deep-learning model such as one-dimensional convolutional network (CN1D) or recurrent neural networks. For a comprehensive survey of ML approaches for this topic, we would refer readers to Rastrollo-Guerrero et al. (2020) and Hellas et al. (2018).

(ii) *Apriori* **attrition prediction**: Students dropping out before course completion is a concerning trend. This is more so in developing nations where very few students finish primary school (Knofczynski, 2017). The outbreak of the COVID-19 pandemic exacerbated the scenario due to indefinite school closures. This led to loss in learning and progress toward providing access to quality education (Moscoviz and Evans, 2022). The causes for dropping out of a course or a degree program can be diverse, but early prediction of it allows administrative staff and educators to intervene. To this end, there have been efforts in using machine learning algorithms to predict attrition.

Massive Open Online Courses (MOOCs): In the context of attrition, special mention must be made of Massive Open Online Courses (MOOCs). While MOOCs promise the democratization of education, one of the biggest concerns with MOOCs is the disparity between the number of students who sign up for a course versus the number of students who actually complete the course—the drop-out rate in MOOCs is significantly high (Hollands and Kazi, 2018; Reich and Ruipérez-Valiente, 2019). Yet in order to make post-secondary and professional education more accessible, MOOCs have become more a practical necessity than an experiment. The COVID-19 pandemic has only emphasized this necessity (Purkayastha and Sinha, 2021). In our literature search phase, we found a sizeable number of contributions in attrition prediction that uses data from MOOC platforms. In this subsection, we will be including those as well as attrition prediction in traditional learning environments.

Early educational data mining methods (Dekker et al., 2009) proposed to predict student drop-out mostly used data sources such as student records (i.e., student demographics, academic, residential, gap year, financial aid information) and administrative records (major administrative changes in education, records of student transfers) to train simple classifiers such as Logistic Regression, Decision Tree, BayesNet, and Random Forest. Selecting an appropriate set of features and designing explainable models has been important as these later inform intervention (Aguiar et al., 2015). To this end, researchers have explored features such as students' prior experiences, motivation and home environment (DeBoer et al., 2013) and student engagement with the course (Aguiar et al., 2014; Ramesh et al., 2014). With the inclusion of an online learning component (particularly relevant for MOOCs), click-stream data and browser information generated allowed researchers to better understand student behavior in an ongoing course. Using historical click-stream data in conjuction with present click-stream data, allowed (Kloft et al., 2014) to effectively predict drop-outs weekly using a simple Support Vector Machine algorithm. This kind of data has also been helpful in understanding the traits indicative of decreased engagement (Sinha et al., 2014), the role of a social cohort structure (Yang et al., 2013) and the sentiment in the student discussion boards and communities (Wen et al., 2014) leading up to student drop-out. He et al. (2015) addresses the concern that weekly prediction of probability of a student dropping out might have wide variance by including smoothing techniques. On the other hand, as resources to intervene might be limited, Lakkaraju et al. (2015) recommends assigning a risk-score per student rather than a binary label. Brooks et al. (2015) considers the level of activity of a student in bins of time during a semester as a binary features (active vs. inactive) and then uses these sequences as n-grams to predict drop-out. Recent developments in

predicting student attrition propose the use of data acquired from disparate sources in addition to more sophisticated algorithms such as deep feed-forward neural networks (Imran et al., 2019) and hybrid logit leaf model (Coussement et al., 2020).

5. Discussion

In this article, we have investigated the involvement of artificial intelligence in the end-to-end educational process. We have highlighted specific research problems both in the planning and in the knowledge delivery phase and reviewed the technological progress in addressing those problems in the past two decades. To the best of our knowledge, such distinction between proactive and reactive phases of education accompanied by a technical deep-dive is an uniqueness of this review.

5.1. Major trends in involvement of AI in the end-to-end education process

The growing interest in AIEd can be inferred from Figures 2, 4 which show how both the count of technical contributions and the count of review articles on the topic have increased over the past two decades. It is to be noted that the number of technical contributions in 2021 and 2022 (assuming our sample of reviewed articles is representative of the population) might have fallen in part due to pandemic-related indefinite school closures and shift to alternate learning models. This triggered a setback on data collection, reporting, and annotation efforts due to a number of factors including lack of direct access to participants, unreliable network connectivity and the necessity of enumerators adopting to new training modes (Wolf et al., 2022). Another important observation from Figure 3 is that AIEd research in most categories focuses heavily on learners in universities, e-learning platforms and MOOCs—work targeting pre-school and K-12 learners is conspicuously absent. A notable exception is research surrounding tutoring aids that has a nearly uniform attention for different target audience groups.

In all categories, to different extents, we see a distinct shift from rule-based and statistical approaches to classical ML to deep learning methods, and from handcrafted features to automatically extracted features. This advancement goes hand-in-hand with the increasingly complex nature of the data being utilized for training AIEd systems. Whereas, earlier approaches used mostly static data (e.g., student records, administrative records, demographic information, surveys, and questionnaires), the use of more sophisticated algorithms necessitated (and in turn benefited from) more real-time and high-volume data (e.g., student-teacher/peer-peer interaction data, click-stream information, web-browsing data). The type of data used by AIEd systems also evolved from mostly tabular records to more text-based and even multi-modal data, spurred on by the emergence of large language models that can handle large quantities of such data.

Even though data-hungry models like deep neural networks have grown in popularity across almost all categories discussed here, AIEd often suffers from the availability of sufficient labeled data to train such systems. This is particularly true for small classes

and new course offerings, or when existing curriculum or tests are changed to incorporate new elements. As a result, another emerging trend in AIEd focuses on using information from resource-rich courses or existing teaching/evaluation content through domain adaptation, transfer learning, few-shot learning, meta learning, etc.

5.2. Impact of COVID-19 pandemic on driving AI research in the frontier of education

COVID-19 pandemic, possibly the most significant social disruptor in recent history, impacted more than 1.5 billion students worldwide (UNESCO, 2022) and is believed to have had far-reaching consequences in the domain of education, possibly even generational setbacks (Tadesse and Muluye, 2020; Dorn et al., 2021; Spector, 2022). As lockdowns and social distancing mandated a hastened transition to fully virtual delivery of educational content, the pandemic era saw an increasing adoption of video conferencing softwares and social media platforms for knowledge delivery, combined with more asynchronous formats of learning. These alternative media of communication were often accompanied by decreasing levels of engagement and satisfaction of learners (Wester et al., 2021; Hollister et al., 2022). There was also a corresponding decrease in practical sessions, labs, and workshops, which are quite critical in some fields of education (Hilburg et al., 2020). However, the pandemic also led to an accelerated adoption of AI-based approaches in education. Pilot studies show that the pandemic led to a significant increase in the usage of AI-based e-learning platforms (Pantelimon et al., 2021). Moreover, a natural by-product of the transition to online learning environments is the generation and logging of more data points from the learning process (Xie et al., 2020) that can be used in AI-based methods to assess and drive student engagement and provide personalized feedback. Online teaching platforms also make it easier to incorporate web-based content, smart interactive elements and asynchronous review sessions to keep students more engaged (Kexin et al., 2020; Pantelimon et al., 2021).

Several recent works have investigated the role of pandemic-driven remote and hybrid instruction in widening gaps in educational achievements by race, poverty level, and gender (Halloran et al., 2021; UNESCO, 2021; Goldhaber et al., 2022). A widespread transition to remote learning necessitates access to proper infrastructure (electricity, internet connectivity, and smart electronic devices that can support video conferencing apps and basic file sharing) as well as resources (learning material, textbooks, educational softwares, etc.), which create barriers for low-income groups (Muñoz-Najar et al., 2021). Even within similar populations, unequal distribution of household chores, income-generating activities, and access to technology-enabled devices affect students of different genders disproportionately (UNESCO, 2021). Moreover, remote learning requires a level of tech-savviness on the part of students and teachers alike, which might be less prevalent in people with learning disabilities. In this context, Garg and Sharma (2020) outlines the different ways AI is used in special need education for development of adaptive and inclusive pedagogies. Salas-Pilco et al. (2022) reviews the different ways

in which AI positively impacts education of minority students, e.g., through facilitating performance/engagement improvement, student retention, student interest in STEM/STEAM fields, etc. Salas-Pilco et al. (2022) also outlines the technological, pedagogical, and socio-cultural barriers for AIEd in inclusive education.

5.3. Existing challenges in adopting artificial intelligence for education

In 2023, artificial intelligence has permeated the lives of people in some aspect or other globally (e.g. chat-bots for customer service, automated credit score analysis, personalized recommendations). At the same time, AI-driven technology for the education sector is gradually becoming a practical necessity globally. The question therefore is, what are the existing barriers in global adoption of AI for education in a safe and inclusive manner—we discuss some of our observations with regards to deploying existing AI driven educational technology at scale.

5.3.1. Lack of concrete legal and ethical guidelines for AIEd research

As pointed out by Pedro et al. (2019), besides most AIEd researchers being concentrated in the technologically advanced parts of the world, most AIEd platforms and applications are owned currently by the private sector. The private investor funded research in big corporations such as Coursera, EdX, IBM, McGraw-Hill, and start-ups like Elsa, Century, Querium have yielded several robust AIEd applications. However, as these platforms are privately owned, there is little transparency and regulations regarding their development and operations. Due to this, there is growing concern on the part of guardians and teaching staff regarding the data accessed by these platforms, privacy, and security of the data stored and explainability of the deployed models. To alleviate this, regulation policies at the international, national, and state levels can help address the concerns of the end users. While many tech-savvy nations have had a head start in this Stirling et al. (2017), drafting general guidelines for AIEd platforms is still very much a nascent concept for most policy makers.

5.3.2. Lack of equitable access to infrastructure hosting AIEd

Education is one of the most important social equalizers (Winthrop, 2018). However, in order to ensure more people have access to quality education, AI-enabled teaching, and studying tools are necessary to reduce the stress on educators and administrative staff (Pedro et al., 2019). The paradox here is that the cost of deploying and operating AIEd tools often alienates communities with limited means thereby widening the gap in access to education. Nye (2015) mentions that access to electricity, internet, data storage, and processing hardware have been barriers in deploying AI-driven platforms. To remove these obstacles, changes must be brought about in local and global levels. While formation of international alliances that invest in infrastructure

development can usher in the technology in developing nations, changes in local policies can expedite the process (Mbangula, 2022).

5.3.3. Lack of skilled personnels to operate AIEd tools in production

Investing in AIEd research and supporting infrastructure alone is not sufficient to ensure long term utility and usage of AI-driven tools for education. Workforce responsible for using these tools on a day-to-day basis must also be brought up to speed. Currently, there is a considerable amount of apprehension, particularly in developing countries, regarding use of AI for education (Shum and Luckin, 2019; Alam, 2021). The main concerns are related to data privacy and security, job security, ethics etc. post adoption of AI in this sector. These concerns in turn have slowed down integration of technology for education. In this context, we must echo (Pedro et al., 2019) in mentioning that while these concerns are relevant and must be addressed, in our review of AIEd research, we have not found any evidence that should invoke consternation in educators and administrative staff. AIEd research as it stands today only augments the role of the teacher, and does not eliminate it. Furthermore, for the foreseeable future, we would need a human in the loop to provide feedback and ensure proper daily usage of these tools.

5.4. Concluding remarks

Through this review, we identified the paradigm shift over the past 20 years in formulating computational models (i.e., choice of algorithms, choice of features etc.) and training them (i.e., choice of data)—we are indeed increasingly leaning toward *sophisticated yet explainable* frameworks. As the scope of this review includes a period of social disruption due to COVID-19 pandemic, it provided us the opportunity to introspect on the utility and the robustness of the proposed technology thus far. To this end, we have discussed the concerns and limitations brought to light by the pandemic and research ideas spawning from that.

With the target of ensuring equitable access to education being set for 2030 by UNGA (United Nations, 2015), one of the inevitable questions arising is: *are we ready to use AI driven ed-tech tools to support educators and students?*. This remains however a question to be answered. Based on our survey, we have observed that while in some parts of the world we have seen great momentum in making AIEd a part and parcel of the education sector, in other parts of the world this progress is stymied by inadequate access to necessary infrastructure and human resources. The ethical and legal implications for large-scale adoption of AI for education is also a topic of active debate (Holmes and Porayska-Pomsta, 2022). The pivotal point at this time is that while there needs to be changes at a socio-economic level to adopt the state of the art AI-driven ed-tech technologies as standard tools for education, the progress made and the ongoing conversations are reasons for positivity.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

A preprint version of this paper is available at: https://arxiv.org/abs/2301.10231 (Mallik and Gangopadhyay, 2023).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2023.1151391/full#supplementary-material

Supplementary section contains the full list of 195 technical articles that have been reviewed in this paper under their respective categories and subcategories.

References

Abdelrahman, G., and Wang, Q. (2019). "Knowledge tracing with sequential key-value memory networks," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris), 175–184. doi: 10.1145/3331184.3331195

Abdelrahman, G., Wang, Q., and Nunes, B. P. (2022). Knowledge tracing: a survey. *ACM Comput. Surveys* 55, 1–37. doi: 10.1145/3569576

Abunasser, B. S., AL-Hiealy, M. R. J., Barhoom, A. M., Almasri, A. R., and Abu-Naser, S. S. (2022). Prediction of instructor performance using machine and deep learning techniques. *Int. J. Adv. Comput. Sci. Appl.* 13, 78–83. doi: 10.14569/IJACSA.2022.0130711

Acampora, G., and Cosma, G. (2015). "A fuzzy-based approach to programming language independent source-code plagiarism detection," in *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (Istanbul), 1–8. doi: 10.1109/FUZZ-IEEE.2015.7337935

Adamson, A., Lamb, A., and December, R. (2014). *Automated Essay Grading.*

Afzal, S., Dhamecha, T. I., Gagnon, P., Nayak, A., Shah, A., Carlstedt-Duke, J., et al. (2020). "AI medical school tutor: modelling and implementation," in *International Conference on Artificial Intelligence in Medicine* (Minneapolis, MN: Springer), 133–145. doi: 10.1007/978-3-030-59137-3_13

Agarwal, M., and Mannem, P. (2011). "Automatic gap-fill question generation from text books," in *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications* (Portland, OR), 56–64.

Agarwal, S. (2020). *Trade-offs between fairness, interpretability, and privacy in machine learning* (Master's thesis). University of Waterloo, Waterloo, ON, Canada.

Aguiar, E., Chawla, N. V., Brockman, J., Ambrose, G. A., and Goodrich, V. (2014). "Engagement vs. performance: using electronic portfolios to predict first semester engineering student retention," in *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (Indianapolis, IN), 103–112. doi: 10.1145/2567574.2567583

Aguiar, E., Lakkaraju, H., Bhanpuri, N., Miller, D., Yuhas, B., and Addison, K. L. (2015). "Who, when, and why: a machine learning approach to prioritizing students at risk of not graduating high school on time," in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (Poughkeepsie, NY), 93–102. doi: 10.1145/2723576.2723619

Ahmad, K., Qadir, J., Al-Fuqaha, A., Iqbal, W., El-Hassan, A., Benhaddou, D., et al. (2020). *Data-Driven Artificial Intelligence in Education: A Comprehensive Review.* EdArXiv.

Ahmad, S. F., Alam, M. M., Rahmat, M. K., Mubarik, M. S., and Hyder, S. I. (2022). Academic and administrative role of artificial intelligence in education. *Sustainability* 14, 1101. doi: 10.3390/su14031101

Ahmed, A. M., Rizaner, A., and Ulusoy, A. H. (2016). Using data mining to predict instructor performance. *Proc. Comput. Sci.* 102, 137–142. doi: 10.1016/j.procs.2016.09.380

Ahmed, U. Z., Gulwani, S., and Karkare, A. (2013). "Automatically generating problems and solutions for natural deduction," in *Twenty-Third International Joint Conference on Artificial Intelligence* (Beijing).

Ahn, J.-w., Tejwani, R., Sundararajan, S., Sipolins, A., O'Hara, S., Paul, A., et al. (2018). "Intelligent virtual reality tutoring system supporting open educational resource access," in *International Conference on Intelligent Tutoring Systems* (Montreal: Springer), 280–286. doi: 10.1007/978-3-319-91464-0_28

Alam, A. (2021). "Possibilities and apprehensions in the landscape of artificial intelligence in education," in *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)* (Nagpur), 1–8. doi: 10.1109/ICCICA52458.2021.9697272

Alfaro, L., Rivera, C., Luna-Urquizo, J., Castañeda, E., and Fialho, F. (2018). Online learning styles identification model, based on the analysis of user interactions within an e-learning platforms, using neural networks and fuzzy logic. *Int. J. Eng. Technol.* 7, 76. doi: 10.14419/ijet.v7i3.13.16328

Alfikri, Z. F., and Purwarianti, A. (2014). Detailed analysis of extrinsic plagiarism detection system using machine learning approach (naive Bayes and SVM). *TELKOMNIKA Indones. J. Electr. Eng.* 12, 7884–7894. doi: 10.11591/telkomnika.v12i11.6652

AlGhamdi, A., Barsheed, A., AlMshjary, H., and AlGhamdi, H. (2020). "A machine learning approach for graduate admission prediction," in *Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing* (Singapore), 155–158. doi: 10.1145/3388818.3393716

Alkhatlan, A., and Kalita, J. (2018). Intelligent tutoring systems: a comprehensive historical survey with recent developments. *arXiv preprint arXiv:1812.09628.* doi: 10.5120/ijca2019918451

AlKhuzaey, S., Grasso, F., Payne, T. R., and Tamma, V. (2021). "A systematic review of data-driven approaches to item difficulty prediction," in *International Conference on Artificial Intelligence in Education* (Utrecht: Springer), 29–41. doi: 10.1007/978-3-030-78292-4_3

Alzahrani, S. (2015). "Arabic plagiarism detection using word correlation in n-grams with k-overlapping approach," in *Proceedings of the Workshops at the 7th Forum for Information Retrieval Evaluation (FIRE)* (Gandhinagar), 123–125.

Arroyo, I., Beal, C., Murray, T., Walles, R., and Woolf, B. (2004). "Wayang outpost: intelligent tutoring for high stakes achievement tests," in *Proceedings of the 7th International Conference on Intelligent Tutoring Systems (ITS2004)* (Maceió), 468–477. doi: 10.1007/978-3-540-30139-4_44

Arroyo, I., Cooper, D. G., Burleson, W., and Woolf, B. P. (2010). "Bayesian networks and linear regression models of students–goals, moods, and emotions," in *Handbook of Educational Data Mining* eds Romero, C., Ventura, S., Pechenizkiy, M., and Baker, R. S. J. d. (Chapman & Hall), 323–338.

Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., and Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *Int. J. Artif. Intell. Educ.* 24, 387–426. doi: 10.1007/s40593-014-0023-y

Ashfaque, M. W., Tharewal, S., Iqhbal, S., and Kayte, C. N. (2020). "A review on techniques, characteristics and approaches of an intelligent tutoring chatbot system," in *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)* (Aurangabad), 258–262. doi: 10.1109/ICSIDEMPC49020.2020.9299583

Assiri, B., Bashraheel, M., and Alsuri, A. (2022). "Improve the accuracy of students admission at universities using machine learning techniques," in *2022 7th International Conference on Data Science and Machine Learning Applications (CDMA)* (Riyadh), 127–132. doi: 10.1109/CDMA54072.2022.00026

Baidoo-Anu, D., and Owusu Ansah, L. (2023). *Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning*. doi: 10.2139/ssrn.4337484

Bajaj, R., and Sharma, V. (2018). Smart education with artificial intelligence based determination of learning styles. *Proc. Comput. Sci.* 132, 834–842. doi: 10.1016/j.procs.2018.05.095

Ball, R., Duhadway, L., Feuz, K., Jensen, J., Rague, B., and Weidman, D. (2019). "Applying machine learning to improve curriculum design," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (Minneapolis, MN), 787–793. doi: 10.1145/3287324.3287430

Bandara, U., and Wijayarathna, G. (2011). A machine learning based tool for source code plagiarism detection. *Int. J. Mach. Learn. Comput.* 1, 337. doi: 10.7763/IJMLC.2011.V1.50

Baral, S., Botelho, A. F., Erickson, J. A., Benachamardi, P., and Heffernan, N. T. (2021). *Improving Automated Scoring of Student Open Responses in Mathematics*. Paris: International Educational Data Mining Society.

Benedetto, L., Cappelli, A., Turrin, R., and Cremonesi, P. (2020a). "Introducing a framework to assess newly created questions with natural language processing," in *International Conference on Artificial Intelligence in Education* (Ifrane: Springer), 43–54. doi: 10.1007/978-3-030-52237-7_4

Benedetto, L., Cappelli, A., Turrin, R., and Cremonesi, P. (2020b). "R2de: a NLP approach to estimating irt parameters of newly generated questions," in *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (Frankfurt), 412–421. doi: 10.1145/3375462.3375517

Benedetto, L., Cremonesi, P., Caines, A., Buttery, P., Cappelli, A., Giussani, A., et al. (2022). A survey on recent approaches to question difficulty estimation from text. *ACM Comput. Surveys* 55, 1–37. doi: 10.1145/3556538

Bernard, J., Chang, T.-W., Popescu, E., and Graf, S. (2015). "Using artificial neural networks to identify learning styles," in *International Conference on Artificial Intelligence in Education* (Madrid: Springer), 541–544. doi: 10.1007/978-3-319-19773-9_57

Blog, O. (2022). *Chatgpt: Optimizing Language Models for Dialogue*.

Bloom, B. S. (1984). The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educ. Res.* 13, 4–16. doi: 10.3102/0013189X013006004

Bogina, V., Hartman, A., Kuflik, T., and Shulner-Tal, A. (2022). Educating software and ai stakeholders about algorithmic fairness, accountability, transparency and ethics. *Int. J. Artif. Intell. Educ.* 32, 808–833. doi: 10.1007/s40593-021-00248-0

Botelho, A. F., Baker, R. S., and Heffernan, N. T. (2017). "Improving sensor-free affect detection using deep learning," in *International Conference on Artificial Intelligence in Education* (Wuhan: Springer), 40–51. doi: 10.1007/978-3-319-61425-0_4

Brooks, C., Thompson, C., and Teasley, S. (2015). "A time series interaction analysis method for building predictive models of learners using log data," in *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (Poughkeepsie, NY), 126–135. doi: 10.1145/2723576.2723581

Bruggink, T. H., and Gambhir, V. (1996). Statistical models for college admission and enrollment: a case study for a selective liberal arts college. *Res. High. Educ.* 37, 221–240. doi: 10.1007/BF01730116

Bryant, J., Heitz, C., Sanghvi, S., and Wagle, D. (2020). *How Artificial Intelligence Will Impact K-12 Teachers*. McKinsey.

Bulman, G., and Fairlie, R. W. (2016). "Technology and education: computers, software, and the internet," in *Handbook of the Economics of Education, Vol.*

5 eds Hanushek, E. A., Machin, S., and Woessmann, L. (Elsevier), 239–280. doi: 10.1016/B978-0-444-63459-7.00005-1

Burkacky, O., Dragon, J., and Lehmann, N. (2022). *The Semiconductor Decade: A Trillion-Dollar Industry*. McKinsey.

Burstein, J., Chodorow, M., and Leacock, C. (2004). Automated essay evaluation: the criterion online writing service. *Ai Mag.* 25, 27. doi: 10.1609/aimag.v25i3.1774

Bydžovská, H. (2016). *A Comparative Analysis of Techniques for Predicting Student Performance*. Raleigh, NC: International Educational Data Mining Society.

Cen, H., Koedinger, K., and Junker, B. (2006). "Learning factors analysis-a general method for cognitive model evaluation and improvement," in *International Conference on Intelligent Tutoring Systems* (Jhongli: Springer), 164–175. doi: 10.1007/11774303_17

Chassignol, M., Khoroshavin, A., Klimova, A., and Bilyatdinova, A. (2018). Artificial intelligence trends in education: a narrative overview. *Proc. Comput. Sci.* 136, 16–24. doi: 10.1016/j.procs.2018.08.233

Chen, L., Chen, P., and Lin, Z. (2020). Artificial intelligence in education: a review. *IEEE Access* 8, 75264–75278. doi: 10.1109/ACCESS.2020.2988510

Chen, S. Y., and Wang, J.-H. (2021). Individual differences and personalized learning: a review and appraisal. *Univers. Access Inform. Soc.* 20, 833–849. doi: 10.1007/s10209-020-00753-4

Chen, X., Zou, D., Xie, H., Cheng, G., and Liu, C. (2022). Two decades of artificial intelligence in education. *Educ. Technol. Soc.* 25, 28–47.

Chen, Y., Wu, L., and Zaki, M. J. (2020). "Reinforcement learning based graph-to-sequence model for natural question generation," in *International Conference on Learning Representations*.

Chui, K. T., Liu, R. W., Zhao, M., and De Pablos, P. O. (2020). Predicting students' performance with school and family tutoring using generative adversarial network-based deep support vector machine. *IEEE Access* 8, 86745–86752. doi: 10.1109/ACCESS.2020.2992869

Contreras, J. O., Hilles, S., and Abubakar, Z. B. (2018). "Automated essay scoring with ontology based on text mining and NLTK tools," in *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)* (Selangor), 1–6. doi: 10.1109/ICSCEE.2018.8538399

Corbett, A. T., and Anderson, J. R. (1994). Knowledge tracing: modeling the acquisition of procedural knowledge. *User Model. User Adapt. Interact.* 4, 253–278. doi: 10.1007/BF01099821

Correia, R., Baptista, J., Eskenazi, M., and Mamede, N. (2012). "Automatic generation of cloze question stems," in *International Conference on Computational Processing of the Portuguese Language* (Coimbra: Springer), 168–178. doi: 10.1007/978-3-642-28885-2_19

Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., and Raes, A. (2020). Predicting student dropout in subscription-based online learning environments: the beneficial impact of the logit leaf model. *Decis. Support Syst.* 135, 113325. doi: 10.1016/j.dss.2020.113325

Cummins, R., Zhang, M., and Briscoe, E. (2016). *Constrained Multi-Task Learning for Automated Essay Scoring*. Association for Computational Linguistics. doi: 10.18653/v1/P16-1075

Dasgupta, T., Naskar, A., Dey, L., and Saha, R. (2018). "Augmenting textual qualitative features in deep convolution recurrent neural network for automatic essay scoring," in *Proceedings of the 5th Workshop on Natural Language Processing Techniques for Educational Applications* (Melbourne), 93–102. doi: 10.18653/v1/W18-3713

Deane, P., and Sheehan, K. (2003). "Automatic item generation via frame semantics: Natural language generation of math word problems," in *Annual Meeting of the National Council of Measurement in Education (ERIC)*.

DeBoer, J., Stump, G. S., Seaton, D., Ho, A., Pritchard, D. E., and Breslow, L. (2013). "Bringing student backgrounds online: MOOC user demographics, site usage, and online learning," in *Educational Data Mining 2013* (Memphis, TN).

DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., et al. (2018). Detecting and addressing frustration in a serious game for military training. *Int. J. Artif. Intell. Educ.* 28, 152–193. doi: 10.1007/s40593-017-0152-1

Dekker, G. W., Pechenizkiy, M., and Vleeshouwers, J. M. (2009). "Predicting students drop out: a case study," in *International Working Group on Educational Data Mining*.

Dey, E. L., and Astin, A. W. (1993). Statistical alternatives for studying college student retention: a comparative analysis of logit, probit, and linear regression. *Res. High. Educ.* 34, 569–581. doi: 10.1007/BF00991920

Dien, T. T., Luu, S. H., Thanh-Hai, N., and Thai-Nghe, N. (2020). Deep learning with data transformation and factor analysis for student performance prediction. *Int. J. Adv. Comput. Sci. Appl.* 11, 711–721. doi: 10.14569/IJACSA.2020.01 10886

Dikli, S. (2006). An overview of automated scoring of essays. *J. Technol. Learn. Assess.* 5.

Dong, N., and Chen, Z. (2020). *The Fourth Education Revolution: Will Artificial Intelligence Liberate or Infantilise Humanity: Buckingham, University of Buckingham*. Springer.

18

Dorn, E., Hancock, B., Sarakatsannis, J., and Viruleg, E. (2021). *COVID-19 and Education: The Lingering Effects of Unfinished Learning*. McKinsey. Available online at: https://www.mckinsey.com/industries/education/our-insights/covid-19-and-education-the-lingering-effects-of-unfinished-learning

Doroudi, S. (2019). *Integrating human and machine intelligence for enhanced curriculum design* (Ph.D. dissertation). Pittsburgh, PA: Air Force Research Laboratory.

Douce, C., Livingstone, D., and Orwell, J. (2005). Automatic test-based assessment of programming: a review. *J. Educ. Resour. Comput.* 5, 4–es. doi: 10.1145/1163405.1163409

Dreyfus, H. L. (1999). Anonymity versus commitment: the dangers of education on the internet. *Ethics Inform. Technol.* 1, 15–20.

Du, X., Shao, J., and Cardie, C. (2017). "Learning to ask: neural question generation for reading comprehension," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vol. 1* (Vancouver), 1342–1352. doi: 10.18653/v1/P17-1123

Duzhin, F., and Gustafsson, A. (2018). Machine learning-based app for self-evaluation of teacher-specific instructional style and tools. *Educ. Sci.* 8:7. doi: 10.3390/educsci8010007

El Guabassi, I., Bousalem, Z., Marah, R., and Qazdar, A. (2021). A recommender system for predicting students' admission to a graduate program using machine learning algorithms. *Int. J. Online Biomed. Engg.* 17, 135–147. doi: 10.3991/ijoe.v17i02.20049

El Mostafa Hambi, F. B. (2020). A new online plagiarism detection system based on deep learning. *Int. J. Adv. Comput. Sci. Appl.* 11, 470–478. doi: 10.14569/IJACSA.2020.0110956

El-Rashidy, M. A., Mohamed, R. G., El-Fishawy, N. A., and Shouman, M. A. (2022). Reliable plagiarism detection system based on deep learning approaches. *Neural Comput. Appl.* 34, 18837–18858. doi: 10.1007/s00521-022-07486-w

Emelianov, V., Gast, N., Gummadi, K. P., and Loiseau, P. (2020). "On fair selection in the presence of implicit variance," in *Proceedings of the 21st ACM Conference on Economics and Computation* (Hungary), 649–675. doi: 10.1145/3391403.3399482

Esparza, G. G., de Luna, A., Zezzatti, A. O., Hernandez, A., Ponce, J., Álvarez, M., et al. (2017). "A sentiment analysis model to analyze students reviews of teacher performance using support vector machines," in *International Symposium on Distributed Computing and Artificial Intelligence* (Porto: Springer), 157–164. doi: 10.1007/978-3-319-62410-5_19

Fahimirad, M., and Kotamjani, S. S. (2018). A review on application of artificial intelligence in teaching and learning in educational contexts. *Int. J. Learn. Dev.* 8, 106–118. doi: 10.5296/ijld.v8i4.14057

Faidhi, J. A., and Robinson, S. K. (1987). An empirical approach for detecting program similarity and plagiarism within a university programming environment. *Comput. Educ.* 11, 11–19. doi: 10.1016/0360-1315(87)90042-X

Fang, J., Zhao, W., and Jia, D. (2019). "Exercise difficulty prediction in online education systems," in *2019 International Conference on Data Mining Workshops (ICDMW)* (Beijing), 311–317. doi: 10.1109/ICDMW.2019.00053

Feenberg, A. (2017). The online education controversy and the future of the university. *Found. Sci.* 22, 363–371. doi: 10.1007/s10699-015-9444-9

Felder, R. M. (1988). Learning and teaching styles in engineering education. *Engg. Educ.* 78, 674–681.

Ferrero, J., Besacier, L., Schwab, D., and Agnès, F. (2017). "Using word embedding for cross-language plagiarism detection," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Vol. 2* (Valencia), 415–421. doi: 10.18653/v1/E17-2066

Finocchiaro, J., Maio, R., Monachou, F., Patro, G. K., Raghavan, M., Stoica, A.-A., et al. (2021). "Bridging machine learning and mechanism design towards algorithmic fairness," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* , 489–503. doi: 10.1145/3442188.3445912

Foltỳnek, T., Meuschke, N., and Gipp, B. (2019). Academic plagiarism detection: a systematic literature review. *ACM Comput. Surveys* 52, 1–42. doi: 10.1145/3345317

Garg, S., and Sharma, S. (2020). Impact of artificial intelligence in special need education to promote inclusive pedagogy. *Int. J. Inform. Educ. Technol.* 10, 523–527. doi: 10.18178/ijiet.2020.10.7.1418

Ghosh, A., Heffernan, N., and Lan, A. S. (2020). "Context-aware attentive knowledge tracing," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (California, CA), 2330–2339. doi: 10.1145/3394486.3403282

Gitinabard, N., Xu, Y., Heckman, S., Barnes, T., and Lynch, C. F. (2019). How widely can prediction models be generalized? An analysis of performance prediction in blended courses. *IEEE Transactions on Learning Technologies.* 12, 184–197. doi: 10.1109/TLT.2019.2911832

Goh, S. L., Kendall, G., and Sabar, N. R. (2019). Simulated annealing with improved reheating and learning for the post enrolment course timetabling problem. *J. Oper. Res. Soc.* 70, 873–888. doi: 10.1080/01605682.2018.1468862

Goldhaber, D., Kane, T. J., McEachin, A., Morton, E., Patterson, T., and Staiger, D. O. (2022). *The Consequences of Remote and Hybrid Instruction During the Pandemic.* Technical report, National Bureau of Economic Research. doi: 10.3386/w30010

Goni, M. O. F., Matin, A., Hasan, T., Siddique, M. A. I., Jyoti, O., and Hasnain, F. M. S. (2020). "Graduate admission chance prediction using deep neural network," in *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)* (Bhubaneswar), 259–262.

Gordon, A., van Lent, M., Van Velsen, M., Carpenter, P., and Jhala, A. (2004). "Branching storylines in virtual reality environments for leadership development," in *Proceedings of the National Conference on Artificial Intelligence* (Menlo Park, CA; Cambridge, MA; London: AAAI Press; MIT Press), 844–851.

Gordon, G., and Breazeal, C. (2015). "Bayesian active learning-based robot tutor for children's word-reading skills," in *Proceedings of the AAAI Conference on Artificial Intelligence* (Austin, TX), Vol. 29. doi: 10.1609/aaai.v29i1.9376

Green, D., Walsh, T., Cohen, P., and Chang, Y.-H. (2011). "Learning a skill-teaching curriculum with dynamic Bayes nets," in *Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 25* (San Francisco, CA), 1648–1654. doi: 10.1609/aaai.v25i2.18855

Grivokostopoulou, F., Hatzilygeroudis, I., and Perikos, I. (2014). Teaching assistance and automatic difficulty estimation in converting first order logic to clause form. *Artif. Intell. Rev.* 42, 347–367. doi: 10.1007/s10462-013-9417-8

Gupta, D., Vani, K., and Singh, C. K. (2014). "Using natural language processing techniques and fuzzy-semantic similarity for automatic external plagiarism detection," in *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (Delhi), 2694–2699. doi: 10.1109/ICACCI.2014.6968314

Gutiérrez, G., Canul-Reich, J., Zezzatti, A. O., Margain, L., and Ponce, J. (2018). Mining: students comments about teacher performance assessment using machine learning algorithms. *Int. J. Combin. Optim. Probl. Inform.* 9, 26.

Haenlein, M., and Kaplan, A. (2019). A brief history of artificial intelligence: on the past, present, and future of artificial intelligence. *Calif. Manage. Rev.* 61, 5–14. doi: 10.1177/0008125619864925

Halloran, C., Jack, R., Okun, J. C., and Oster, E. (2021). *Pandemic Schooling Mode and Student Test Scores: Evidence From US States.* Technical report, National Bureau of Economic Research. doi: 10.3386/w29497

Hänig, C., Remus, R., and De La Puente, X. (2015). "EXB themis: extensive feature extraction from word alignments for semantic textual similarity," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)* (Denver, TX), 264–268. doi: 10.18653/v1/S15-2046

Harley, J. M., Lajoie, S. P., Frasson, C., and Hall, N. C. (2017). Developing emotion-aware, advanced learning technologies: a taxonomy of approaches and features. *Int. J. Artif. Intell. Educ.* 27, 268–297. doi: 10.1007/s40593-016-0126-8

He, J., Bailey, J., Rubinstein, B., and Zhang, R. (2015). "Identifying at-risk students in massive open online courses," in *Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 29* (Austin, TX). doi: 10.1609/aaai.v29i1.9471

Heilman, M. (2011). *Automatic factual question generation from text* (Ph.D. thesis). Carnegie Mellon University, Pittsburgh, PA, United States.

Heilman, M., and Smith, N. A. (2010). "Good question! Statistical ranking for question generation," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (Los Angeles, CA), 609–617.

Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V. V., Gutica, M., Hynninen, T., et al. (2018). "Predicting academic performance: a systematic literature review," in *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (Larnaca), 175–199. doi: 10.1145/3293881.3295783

Herzog, S. (2006). Estimating student retention and degree-completion time: decision trees and neural networks vis-à-vis regression. *New Direct. Instit. Res.* 131, 17–33. doi: 10.1002/ir.185

Hilburg, R., Patel, N., Ambruso, S., Biewald, M. A., and Farouk, S. S. (2020). Medical education during the coronavirus disease-2019 pandemic: learning from a distance. *Adv. Chron. Kidney Dis.* 27, 412–417. doi: 10.1053/j.ackd.2020.05.017

Hockly, N. (2019). Automated writing evaluation. *ELT. J.* 73, 82–88. doi: 10.1093/elt/ccy044

Hollands, F., and Kazi, A. (2018). *Benefits and Costs of MOOC-Based Alternative Credentials.* Center for Benefit-Cost Studies of Education.

Hollister, B., Nair, P., Hill-Lindsay, S., and Chukoskie, L. (2022). Engagement in online learning: student attitudes and behavior during COVID-19. *Front. Educ.* 7, 851019. doi: 10.3389/feduc.2022.851019

Holmes, W., and Porayska-Pomsta, K. (2022). *The Ethics of Artificial Intelligence in Education: Practices, Challenges, and Debates.* Taylor & Francis. doi: 10.4324/9780429329067

Holmes, W., and Tuomi, I. (2022). State of the art and practice in AI in education. *Eur. J. Educ.* 57, 542–570. doi: 10.1111/ejed.12533

Honey, P., and Mumford, A. (1986). *The Manual of Learning Styles.*

Hu, J. (2021). Teaching evaluation system by use of machine learning and artificial intelligence methods. *Int. J. Emerg. Technol. Learn.* 16, 87–101. doi: 10.3991/ijet.v16i05.20299

Huang, J., Piech, C., Nguyen, A., and Guibas, L. (2013). "Syntactic and functional variability of a million code submissions in a machine learning MOOC," in *AIED 2013 Workshops Proceedings, Vol. 25* (Memphis, TN).

Huang, J., Saleh, S., and Liu, Y. (2021). A review on artificial intelligence in education. *Acad. J. Interdisc. Stud.* 10, 206. doi: 10.36941/ajis-2021-0077

Huang, Y., Huang, W., Tong, S., Huang, Z., Liu, Q., Chen, E., et al. (2021). "Stan: adversarial network for cross-domain question difficulty prediction," in *2021 IEEE International Conference on Data Mining (ICDM)* (Auckland), 220–229. doi: 10.1109/ICDM51629.2021.00032

Huang, Y.-T., Tseng, Y.-M., Sun, Y. S., and Chen, M. C. (2014). "Tedquiz: automatic quiz generation for ted talks video clips to assess listening comprehension," in *2014 IEEE 14Th International Conference on Advanced Learning Technologies* (Athens), 350–354. doi: 10.1109/ICALT.2014.105

Huang, Z., Liu, Q., Chen, E., Zhao, H., Gao, M., Wei, S., et al. (2017). "Question difficulty prediction for reading problems in standard tests," in *Thirty-First AAAI Conference on Artificial Intelligence* (San Francisco, CA). doi: 10.1609/aaai.v31i1.10740

Humble, N., and Mozelius, P. (2019). "Artificial intelligence in education–a promise, a threat or a hype," in *Proceedings of the European Conference on the Impact of Artificial Intelligence and Robotics* (Oxford), 149–156.

Hürlimann, M., Weck, B., van den Berg, E., Suster, S., and Nissim, M. (2015). "Glad: Groningen lightweight authorship detection," in *CLEF (Working Notes)* (Toulouse).

Hwang, G.-J., Xie, H., Wah, B. W., and Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Comput. Educ. Artif. Intell.* 1, 10001. doi: 10.1016/j.caeai.2020.100001

Idris, N., Yusof, N., and Saad, P. (2009). Adaptive course sequencing for personalization of learning path using neural network. *Int. J. Adv. Soft Comput. Appl.* 1, 49–61.

Imran, A. S., Dalipi, F., and Kastrati, Z. (2019). "Predicting student dropout in a MOOC: an evaluation of a deep neural network model," in *Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence* (Bali), 190–195. doi: 10.1145/3330482.3330514

Indurthi, S. R., Raghu, D., Khapra, M. M., and Joshi, S. (2017). "Generating natural language question-answer pairs from a knowledge graph using a RNN based question generation model," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Vol. 1* (Valencia), 376–385.

Islam, M. Z., Ali, R., Haider, A., Islam, M. Z., and Kim, H. S. (2021). Pakes: a reinforcement learning-based personalized adaptability knowledge extraction strategy for adaptive learning systems. *IEEE Access* 9, 155123–155137. doi: 10.1109/ACCESS.2021.3128578

Jackson, G. T., and McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *J. Educ. Psychol.* 105, 1036. doi: 10.1037/a0032580

Jamison, J. (2017). "Applying machine learning to predict Davidson college's admissions yield," in *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education* (Seattle, WA), 765–766. doi: 10.1145/3017680.3022468

Ji, J.-H., Woo, G., and Cho, H.-G. (2007). "A source code linearization technique for detecting plagiarized programs," in *Proceedings of the 12th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education* (Dundee), 73–77. doi: 10.1145/1269900.1268807

Jiang, S., Williams, A., Schenke, K., Warschauer, M., and O'dowd, D. (2014). "Predicting MOOC performance with week 1 behavior," in *Educational Data Mining 2014* (London).

Jouault, C., and Seta, K. (2013). "Building a semantic open learning space with adaptive question generation support," in *Proceedings of the 21st International Conference on Computers in Education* (Bali), 41–50.

Kalady, S., Elikkottil, A., and Das, R. (2010). "Natural language question generation using syntax and keywords," in *Proceedings of QG2010: The Third Workshop on Question Generation, Vol. 2* (Pittsburgh, PA), 5–14.

Katta, J. Y. B. (2018). *Machine learning for source-code plagiarism detection* (Ph.D. thesis). International Institute of Information Technology Hyderabad.

Ke, Z., Inamdar, H., Lin, H., and Ng, V. (2019). "Give me more feedback ii: annotating thesis strength and related attributes in student essays," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (Florence), 3994–4004.

Keller, S. U. (2021). Automatic generation of word problems for academic education via natural language processing (nlp). *arXiv preprint arXiv:2109.13123.* doi: 10.48550/arXiv.2109.13123

Kenekayoro, P. (2019). Incorporating machine learning to evaluate solutions to the university course timetabling problem. *Covenant J. Inform. Commun. Technol.* 7:18–35. doi: 10.48550/arXiv.2010.00826

Kessels, J. (1999). "A relational approach to curriculum design," in *Design Approaches and Tools in Education and Training* (Springer), 59–70. doi: 10.1007/978-94-011-4255-7_5

Kexin, L., Yi, Q., Xiaoou, S., and Yan, L. (2020). "Future education trend learned from the COVID-19 pandemic: take artificial intelligence online course as an example," in *2020 International Conference on Artificial Intelligence and Education (ICAIE)* (Tianjin), 108–111. doi: 10.1109/ICAIE50891.2020.00032

Khajah, M., Wing, R., Lindsey, R. V., and Mozer, M. (2014). "Integrating latent-factor and knowledge-tracing models to predict individual differences in learning," in *EDM* (London), 99–106.

Kim, J., and Shaw, E. (2009). "Pedagogical discourse: connecting students to past discussions and peer mentors within an online discussion board," in *Twenty-First IAAI Conference* (Pasadena, MD).

Kim, Y., Lee, H., Shin, J., and Jung, K. (2019). "Improving neural question generation using answer separation," in *Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33* (Honolulu, HI), 6602–6609. doi: 10.1609/aaai.v33i01.33016602

Kloft, M., Stiehler, F., Zheng, Z., and Pinkwart, N. (2014). "Predicting MOOC dropout over weeks using machine learning methods," in *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs* (Doha), 60–65. doi: 10.3115/v1/W14-4111

Knofczynski, A. (2017). *Why Global Drop-Out Rates Aren't Improving.* The Borgen Project.

Kolb, D. A. (1976). *Learning Style Inventory: Technical Manual.* Boston, MA: McBer.

Koncel-Kedziorski, R., Konstas, I., Zettlemoyer, L., and Hajishirzi, H. (2016). "A theme-rewriting approach for generating algebra word problems," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (Austin, TX), 1617–28. doi: 10.18653/v1/D16-1168

Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., et al. (2015). "A machine learning framework to identify students at risk of adverse academic outcomes," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Sydney), 1909–1918. doi: 10.1145/2783258.2788620

Lameras, P., and Arnab, S. (2021). Power to the teachers: an exploratory review on artificial intelligence in education. *nformation* 13, 14. doi: 10.3390/info13010014

Lan, A. S., Vats, D., Waters, A. E., and Baraniuk, R. G. (2015). "Mathematical language processing: automatic grading and feedback for open response mathematical questions," in *Proceedings of the Second (2015) ACM Conference on Learning@ scale* (Vancouver), 167–176. doi: 10.1145/2724660.2724664

Lange, R. C., and Mancoridis, S. (2007). "Using code metric histograms and genetic algorithms to perform author identification for software forensics," in *Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation* (London), 2082–2089. doi: 10.1145/1276958.1277364

Larabi-Marie-Sainte, S., Jan, R., Al-Matouq, A., and Alabduhadi, S. (2021). The impact of timetable on student's absences and performance. *PLoS ONE* 16, e0253256. doi: 10.1371/journal.pone.0253256

Li, H., Wang, S., Liang, J., Huang, S., and Xu, B. (2009). "High performance automatic mispronunciation detection method based on neural network and trap features," in *Tenth Annual Conference of the International Speech Communication Association* (Brighton).

Li, Q., and Kim, J. (2021). A deep learning-based course recommender system for sustainable development in education. *Appl. Sci.* 11, 8993. doi: 10.3390/app11198993

Li, W., Li, K., Siniscalchi, S. M., Chen, N. F., and Lee, C.-H. (2016). "Detecting mispronunciations of l2 learners and providing corrective feedback using knowledge-guided and data-driven decision trees," in *Interspeech* (San Francisco, CA), 3127–3131.

Li, X., Chen, M., and Nie, J.-Y. (2020). SEDNN: shared and enhanced deep neural network model for cross-prompt automated essay scoring. *Knowl. Based Syst.* 210, 106491. doi: 10.1016/j.knosys.2020.106491

Lin, C. F., Yeh, Y.-C., Hung, Y. H., and Chang, R. I. (2013). Data mining for providing a personalized learning path in creativity: an application of decision trees. *Comput. Educ.* 68, 199–210. doi: 10.1016/j.compedu.2013.05.009

Lin, L.-H., Chang, T.-H., and Hsu, F.-Y. (2019). "Automated prediction of item difficulty in reading comprehension using long short-term memory," in *2019 International Conference on Asian Language Processing (IALP)* (Shanghai), 132–135. doi: 10.1109/IALP48816.2019.9037716

Lin, M.-H., Chen, H.-G., and Liu, K. S. (2017). A study of the effects of digital learning on learning motivation and learning outcome. *Eur. J. Math. Sci. Technol. Educ.* 13, 3553–3564. doi: 10.12973/eurasia.2017.00744a

Lindberg, D., Popowich, F., Nesbit, J., and Winne, P. (2013). "Generating natural language questions to support learning on-line," in *Proceedings of the 14th European Workshop on Natural Language Generation* (Sofia), 105–114.

Liu, Q., Huang, Z., Yin, Y., Chen, E., Xiong, H., Su, Y., et al. (2019). EKT: exercise-aware knowledge tracing for student performance prediction. *IEEE Trans. Knowl. Data Eng.* 33, 100–115. doi: 10.1109/TKDE.2019.2924374

Liu, T., Fang, Q., Ding, W., Li, H., Wu, Z., and Liu, Z. (2021). "Mathematical word problem generation from commonsense knowledge graph and equations," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 4225–4240. doi: 10.18653/v1/2021.emnlp-main.348

Lo, J.-J., and Shu, P.-C. (2005). Identification of learning styles online by observing learners' browsing behaviour through a neural network. *Br. J. Educ. Technol.* 36, 43–55. doi: 10.1111/j.1467-8535.2005.00437.x

Lund, B. D., and Wang, T. (2023). Chatting about chatGPT: how may AI and GPT impact academia and libraries? *Library Hi Tech News.* doi: 10.1108/LHTN-01-2023-0009

Majumder, M., and Saha, S. K. (2015). "A system for generating multiple choice questions: with a novel approach for sentence selection," in *Proceedings of the 2nd Workshop on Natural Language Processing Techniques for Educational Applications* (Beijing), 64–72. doi: 10.18653/v1/W15-4410

Malik, G., Tayal, D. K., and Vij, S. (2019). "An analysis of the role of artificial intelligence in education and teaching," in *Recent Findings in Intelligent Computing Techniques* eds Sa, P. K., Bakshi, S., Hatzilygeroudis, I. K., and Sahoo, M. N. (Springer), 407–417. doi: 10.1007/978-981-10-8639-7_42

Mallik, S., and Gangopadhyay, A. (2023). Proactive and reactive engagement of artificial intelligence methods for education: a review. *arXiv preprint arXiv:2301.10231.*

Manahi, M. S. (2021). *A deep learning framework for the defection of source code plagiarism using siamese network and embedding models* (Master's thesis). Kulliyyah of Information and Communication Technology, Kuala Lumpur, Malaysia. doi: 10.1007/978-981-16-8515-6_31

Marcinkowski, F., Kieslich, K., Starke, C., and Lünich, M. (2020). "Implications of AI (un-) fairness in higher education admissions: the effects of perceived AI (un-) fairness on exit, voice and organizational reputation," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona), 122–130. doi: 10.1145/3351095.3372867

Mardikyan, S., and Badur, B. (2011). Analyzing teaching performance of instructors using data mining techniques. *Inform. Educ.* 10, 245–257. doi: 10.15388/infedu.2011.17

Mbangula, D. K. (2022). "Adopting of artificial intelligence and development in developing countries: perspective of economic transformation," in *Handbook of Research on Connecting Philosophy, Media, and Development in Developing Countries* eds Okocha, D. O., Onobe, M. J., and Alike, M. N. (IGI Global), 276–288. doi: 10.4018/978-1-6684-4107-7.ch018

Mei, X. Y., Aas, E., and Medgard, M. (2019). Teachers' use of digital learning tool for teaching in higher education: exploring teaching practice and sharing culture. *J. Appl. Res. High. Educ.* 11, 522–537. doi: 10.1108/JARHE-10-2018-0202

Mendis, C., Lahiru, D., Pamudika, N., Madushanka, S., Ranathunga, S., and Dias, G. (2017). "Automatic assessment of student answers for geometric theorem proving questions," in *2017 Moratuwa Engineering Research Conference (MERCon)* (Moratuwa), 413–418. doi: 10.1109/MERCon.2017.7980520

Meuschke, N., Siebeck, N., Schubotz, M., and Gipp, B. (2017). "Analyzing semantic concept patterns to detect academic plagiarism," in *Proceedings of the 6th International Workshop on Mining Scientific Publications* (Toronto), 46–53. doi: 10.1145/3127526.3127535

Mirchi, N., Bissonnette, V., Yilmaz, R., Ledwos, N., Winkler-Schwartz, A., and Del Maestro, R. F. (2020). The virtual operative assistant: an explainable artificial intelligence tool for simulation-based training in surgery and medicine. *PLoS ONE* 15, e0229596. doi: 10.1371/journal.pone.0229596

Mokbel, B., Gross, S., Paassen, B., Pinkwart, N., and Hammer, B. (2013). "Domain-independent proximity measures in intelligent tutoring systems," in *Educational Data Mining 2013* (Memphis, TN).

Moore, J. S. (1998). An expert system approach to graduate school admission decisions and academic performance prediction. *Omega* 26, 659–670. doi: 10.1016/S0305-0483(98)00008-5

Moscoviz, L., and Evans, D. (2022). *Learning Loss and Student Dropouts During the Covid-19 Pandemic: A Review of the Evidence Two Years After Schools Shut Down.* Center for Global Development.

Mostafazadeh, N., Misra, I., Devlin, J., Mitchell, M., He, X., and Vanderwende, L. (2016). "Generating natural questions about an image," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Vol. 1* (Berlin), 1802–1813. doi: 10.18653/v1/P16-1170

Mostow, J., and Chen, W. (2009). "Generating instruction automatically for the reading strategy of self-questioning," in *AIED* (Brighton), 465–472.

Mota, J. (2008). "Using learning styles and neural networks as an approach to elearning content and layout adaptation," in *Doctoral Symposium on Informatics Engineering* (Porto).

Mothe, J., and Tanguy, L. (2005). "Linguistic features to predict query difficulty," in *ACM Conference on Research and Development in Information Retrieval, SIGIR, Predicting Query Difficulty-Methods and Applications Workshop* (Salvador), 7–10.

Movellan, J., Eckhardt, M., Virnes, M., and Rodriguez, A. (2009). "Sociable robot improves toddler vocabulary skills," in *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction* (La Jolla, CA), 307–308. doi: 10.1145/1514095.1514189

Mridha, K., Jha, S., Shah, B., Damodharan, P., Ghosh, A., and Shaw, R. N. (2022). "Machine learning algorithms for predicting the graduation admission," in *International Conference on Electrical and Electronics Engineering* (Greater Noida: Springer), 618–637. doi: 10.1007/978-981-19-1677-9_55

Muñoz-Najar, A., Gilberto, A., Hasan, A., Cobo, C., Azevedo, J. P., and Akmal, M. (2021). Remote learning during COVID-19: Lessons from today, principles for tomorrow. *World Bank.* doi: 10.1596/36665

Nadeem, F., Nguyen, H., Liu, Y., and Ostendorf, M. (2019). "Automated essay scoring with discourse-aware neural models," in *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications* (Florence), 484–493. doi: 10.18653/v1/W19-4450

Nagatani, K., Zhang, Q., Sato, M., Chen, Y.-Y., Chen, F., and Ohkuma, T. (2019). "Augmenting knowledge tracing by considering forgetting behavior," in *The World Wide Web Conference* (San Francisco, CA), 3101–3107. doi: 10.1145/3308558.3313565

Nakagawa, H., Iwasawa, Y., and Matsuo, Y. (2019). "Graph-based knowledge tracing: modeling student proficiency using graph neural network," in *2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (Thessaloniki), 156–163. doi: 10.1145/3350546.3352513

Namatherdhala, B., Mazher, N., and Sriram, G. K. (2022). A comprehensive overview of artificial intelligence tends in education. *Int. Res. J. Modern. Eng. Technol. Sci.* 4.

Nghe, N. T., Janecek, P., and Haddawy, P. (2007). "A comparative analysis of techniques for predicting academic performance," in *2007 37th Annual Frontiers in Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports* (Milwaukee, WI).

Nguyen, T., Rosenberg, M., Song, X., Gao, J., Tiwary, S., Majumder, R., et al. (2016). "MS MARCO: a human generated machine reading comprehension dataset," in *CoCo@ NIPs* (Barcelona).

Nye, B. D. (2015). Intelligent tutoring systems by and for the developing world: a review of trends and approaches for educational technology in a global context. *Int. J. Artif. Intell. Educ.* 25, 177–203. doi: 10.1007/s40593-014-0028-6

Obit, J. H., Landa-Silva, D., Sevaux, M., and Ouelhadj, D. (2011). "Non-linear great deluge with reinforcement learning for university course timetabling," in *Metaheuristics-Intelligent Decision Making, Series Operations Research/Computer Science Interfaces* (Springer), 1–19.

Ohmann, T., and Rahal, I. (2015). Efficient clustering-based source code plagiarism detection using piy. *Knowl. Inform. Syst.* 43, 445–472. doi: 10.1007/s10115-014-0742-2

Okoye, K., Arrona-Palacios, A., Camacho-Zuñiga, C., Achem, J. A. G., Escamilla, J., and Hosseini, S. (2022). Towards teaching analytics: a contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification. *Educ. Inform. Technol.* 27, 3891–3933. doi: 10.1007/s10639-021-10751-5

Olney, A. M., D'Mello, S., Person, N., Cade, W., Hays, P., Williams, C., et al. (2012). "Guru: a computer tutor that models expert human tutors," in *International Conference on Intelligent Tutoring Systems* (Chania: Springer), 256–261. doi: 10.1007/978-3-642-30950-2_32

Onan, A. (2020). Mining opinions from instructor evaluation reviews: a deep learning approach. *Comput. Appl. Eng. Educ.* 28, 117–138. doi: 10.1002/cae.22179

Ouyang, F., and Jiao, P. (2021). Artificial intelligence in education: the three paradigms. *Comput. Educ. Artif. Intell.* 2, 100020. doi: 10.1016/j.caeai.2021.100020

Ouyang, F., Zheng, L., and Jiao, P. (2022). Artificial intelligence in online higher education: a systematic review of empirical research from 2011 to 2020. *Educ. Inform. Technol.* 1–33. doi: 10.1007/s10639-022-10925-9

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., et al. (2022). Training language models to follow instructions with human feedback. *Adv. Neural Inform. Process. Syst.* 35, 27730–27744.

Özcan, E., Misir, M., Ochoa, G., and Burke, E. K. (2012). "A reinforcement learning: great-deluge hyper-heuristic for examination timetabling," in *Modeling, Analysis, and Applications in Metaheuristic Computing: Advancements and Trends* (IGI Global), 34–55. doi: 10.4018/978-1-4666-0270-0.ch003

Pan, L., Lei, W., Chua, T.-S., and Kan, M.-Y. (2019). Recent advances in neural question generation. *arXiv preprint arXiv:1905.08949.* doi: 10.48550/arXiv.1905.08949

Pande, C., Witschel, H. F., Martin, A., and Montecchiari, D. (2021). "Hybrid conversational AI for intelligent tutoring systems," in *AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering.* (Virtual).

Pandey, S., and Karypis, G. (2019). "A self-attentive model for knowledge tracing," in *12th International Conference on Educational Data Mining, EDM 2019* (Montreal: International Educational Data Mining Society), 384–389.

Pantelimon, F.-V., Bologa, R., Toma, A., and Posedaru, B.-S. (2021). The evolution of AI-driven educational systems during the COVID-19 pandemic. *Sustainability* 13, 13501. doi: 10.3390/su132313501

Pask, G. (1976). Styles and strategies of learning. *Br. J. Educ. Psychol.* 46, 128–148. doi: 10.1111/j.2044-8279.1976.tb02305.x

Pavlik, P. I., Jr, and Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: an activation-based model of the spacing effect. *Cogn. Sci.* 29, 559–586. doi: 10.1207/s15516709cog0000_14

Pavlik, P. I. Jr, Cen, H., and Koedinger, K. R. (2009). "Performance factors analysis– A new alternative to knowledge tracing," in *Proceedings of the 14th International Conference of Artificial Intelligence in Education* (Brighton).

Pedro, F., Subosa, M., Rivas, A., and Valverde, P. (2019). *Artificial intelligence in Education: Challenges and Opportunities for Sustainable Development*. UNESCO.

Pereira, J. (2016). "Leveraging chatbots to improve self-guided learning through conversational quizzes," in *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality* (Salamanca), 911–918. doi: 10.1145/3012430.3012625

Perikos, I., Grivokostopoulou, F., Kovas, K., and Hatzilygeroudis, I. (2016). Automatic estimation of exercises' difficulty levels in a tutoring system for teaching the conversion of natural language into first-order logic. *Expert Syst.* 33, 569–580. doi: 10.1111/exsy.12182

Persing, I., Davis, A., and Ng, V. (2010). "Modeling organization in student essays," in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* (Cambridge, MA), 229–239.

Persing, I., and Ng, V. (2013). "Modeling thesis clarity in student essays," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Vol. 1* (Sofia), 260–269. doi: 10.3115/v1/P14-1144

Persing, I., and Ng, V. (2014). "Modeling prompt adherence in student essays," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Vol. 1* (Baltimore, MA), 1534–1543.

Persing, I., and Ng, V. (2015). "Modeling argument strength in student essays," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Vol. 1* (Beijing), 543–552. doi: 10.3115/v1/P15-1053

Pertile, S. d. L., Moreira, V. P., and Rosso, P. (2016). Comparing and combining content-and citation-based approaches for plagiarism detection. *J. Assoc. Inform. Sci. Technol.* 67, 2511–2526. doi: 10.1002/asi.23593

Phandi, P., Chai, K. M. A., and Ng, H. T. (2015). "Flexible domain adaptation for automated essay scoring using correlated linear regression," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (Lisbon), 431–439. doi: 10.18653/v1/D15-1049

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., et al. (2015a). "Deep knowledge tracing," in *Advances in Neural Information Processing Systems, Vol. 28* (Montreal).

Piech, C., Huang, J., Nguyen, A., Phulsuksombati, M., Sahami, M., and Guibas, L. (2015b). "Learning program embeddings to propagate feedback on student code," in *International conference on machine Learning* (Lille), 1093–1102.

Polozov, O., O'Rourke, E., Smith, A. M., Zettlemoyer, L., Gulwani, S., and Popović, Z. (2015). "Personalized mathematical word problem generation," in *Twenty-Fourth International Joint Conference on Artificial Intelligence* (Buenos Aires).

Pu, Y., Wang, C., and Wu, W. (2020). "A deep reinforcement learning framework for instructional sequencing," in *2020 IEEE International Conference on Big Data (Big Data)* (Virtual), 5201–5208.

Purkayastha, N., and Sinha, M. K. (2021). "Unstoppable study with MOOCs during COVID 19 pandemic: a study," in *Library Philosophy and Practice* (Lincoln, NE: University of Nebraska), 1–12. doi: 10.2139/ssrn.3978886

Qiu, Z., Wu, X., and Fan, W. (2019). "Question difficulty prediction for multiple choice problems in medical exams," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (Beijing), 139–148. doi: 10.1145/3357384.3358013

Rajpurkar, P., Zhang, J., Lopyrev, K., and Liang, P. (2016). Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*. doi: 10.18653/v1/D16-1264

Ramesh, A., Goldwasser, D., Huang, B., Daumé III, H., and Getoor, L. (2013). "Modeling learner engagement in MOOCs using probabilistic soft logic," in *NIPS Workshop on Data Driven Education, Vol. 21* (Lake Tahoe, NV), 62.

Ramesh, A., Goldwasser, D., Huang, B., Daume III, H., and Getoor, L. (2014). "Learning latent engagement patterns of students in online courses," in *Twenty-Eighth AAAI Conference on Artificial Intelligence* (Quebec). doi: 10.1609/aaai.v28i1.8920

Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., and Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: a review. *Appl. Sci.* 10, 1042. doi: 10.3390/app10031042

Rawatlal, R. (2017). "Application of machine learning to curriculum design analysis," in *2017 Computing Conference* (London), 1143–1151. doi: 10.1109/SAI.2017.8252234

Reddy, S., Levine, S., and Dragan, A. (2017). "Accelerating human learning with deep reinforcement learning," in *NIPS'17 Workshop: Teaching Machines, Robots, and Humans* (Long Beach, CA), 5–9. doi: 10.15607/RSS.2018. XIV.005

Reich, J., and Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science* 363, 130–131. doi: 10.1126/science.aav7958

Salas-Pilco, S. Z., Xiao, K., and Oshima, J. (2022). Artificial intelligence and new technologies in inclusive education for minority students: a systematic review. *Sustainability* 14, 13572. doi: 10.3390/su142013572

San Pedro, M. O. Z., Baker, R. S., Gowda, S. M., and Heffernan, N. T. (2013). "Towards an understanding of affect and knowledge from student interaction with an intelligent tutoring system," in *International Conference on Artificial Intelligence in Education* (Memphis, TN: Springer), 41–50. doi: 10.1007/978-3-642-39112-5_5

Schiff, D. (2021). Out of the laboratory and into the classroom: the future of artificial intelligence in education. *AI Soc.* 36, 331–348. doi: 10.1007/s00146-020-01033-8

Shashidhar, V., Pandey, N., and Aggarwal, V. (2015). "Automatic spontaneous speech grading: a novel feature derivation technique using the crowd," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Vol. 1* (Beijing), 1085–1094. doi: 10.3115/v1/P15-1105

Shen, S., Liu, Q., Chen, E., Huang, Z., Huang, W., Yin, Y., et al. (2021). "Learning process-consistent knowledge tracing," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (Singapore), 1452–1460. doi: 10.1145/3447548.3467237

Shermis, M. D., and Burstein, J. C. (2003). *Automated Essay Scoring: A Cross-Disciplinary Perspective*. Routledge. doi: 10.4324/9781410606860

Shi, Z. R., Wang, C., and Fang, F. (2020). Artificial intelligence for social good: a survey. *arXiv preprint arXiv:2001.01818*. doi: 10.48550/arXiv.2001.01818

Shum, S. J. B., and Luckin, R. (2019). Learning analytics and AI: politics, pedagogy and practices. *Br. J. Educ. Technol.* 50, 2785–2793. doi: 10.1111/bjet.12880

Singh, G., Srikant, S., and Aggarwal, V. (2016). "Question independent grading using machine learning: the case of computer program grading," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, CA), 263–272. doi: 10.1145/2939672.2939696

Sinha, T., Li, N., Jermann, P., and Dillenbourg, P. (2014). "Capturing 'attrition intensifying' structural traits from didactic interaction sequences of MOOC learners," in *EMNLP 2014* (Doha), 42. doi: 10.3115/v1/W14-4108

Soleman, S., and Purwarianti, A. (2014). "Experiments on the Indonesian plagiarism detection using latent semantic analysis," in *2014 2nd International Conference on Information and Communication Technology (ICoICT)* (Bandung), 413–418. doi: 10.1109/ICoICT.2014.6914098

Somasundaram, M., Latha, P., and Pandian, S. S. (2020). Curriculum design using artificial intelligence (AI) back propagation method. *Proc. Comput. Sci.* 172, 134–138. doi: 10.1016/j.procs.2020.05.020

Song, W., Zhang, K., Fu, R., Liu, L., Liu, T., and Cheng, M. (2020). "Multi-stage pre-training for automated Chinese essay scoring," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Virtual), 6723–6733. doi: 10.18653/v1/2020.emnlp-main.546

Spector, C. (2022). *New Research Details the Pandemic's Variable Impact on U.S. School Districts*. Stanford News. Available online at: https://news.stanford.edu/2022/10/28/new-research-details-pandemics-impact-u-s-school-districts

Srikant, S., and Aggarwal, V. (2014). "A system to grade computer programming skills using machine learning," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY), 1887–1896. doi: 10.1145/2623330.2623377

Stiller, J., Hartmann, S., Mathesius, S., Straube, P., Tiemann, R., Nordmeier, V., et al. (2016). Assessing scientific reasoning: a comprehensive evaluation of item features that affect item difficulty. *Assess. Eval. High. Educ.* 41, 721–732. doi: 10.1080/02602938.2016.1164830

Stirling, R., Miller, H., and Martinho-Truswell, E. (2017). Government ai readiness index. *Korea* 4, 7812407479.

Su, J., and Yang, W. (2022). Artificial intelligence in early childhood education: a scoping review. *Comput. Educ.* 2022, 100049. doi: 10.1016/j.caeai.2022.100049

Su, Y., Liu, Q., Liu, Q., Huang, Z., Yin, Y., Chen, E., et al. (2018). "Exercise-enhanced sequential modeling for student performance prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32* (New Orleans, LA). doi: 10.1609/aaai.v32i1.11864

Sultan, M. A., Bethard, S., and Sumner, T. (2014). "Dls @ cu: Sentence similarity from word alignment," in *SemEval@ COLING* (Dublin), 241–246. doi: 10.3115/v1/S14-2039

Tadesse, S., and Muluye, W. (2020). The impact of COVID-19 pandemic on education system in developing countries: a review. *Open J. Soc. Sci.* 8, 159–170. doi: 10.4236/jss.2020.810011

Taghipour, K., and Ng, H. T. (2016). "A neural approach to automated essay scoring," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (Austin, TX), 1882–1891. doi: 10.18653/v1/D16-1193

Tamhane, A., Ikbal, S., Sengupta, B., Duggirala, M., and Appleton, J. (2014). "Predicting student risks through longitudinal analysis," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY), 1544–1552. doi: 10.1145/2623330.2623355

Tan, J. S., Goh, S. L., Kendall, G., and Sabar, N. R. (2021). A survey of the state-of-the-art of optimisation methodologies in school timetabling problems. *Expert Syst. Appl.* 165, 113943. doi: 10.1016/j.eswa.2020.113943

Tan, S., Doshi-Velez, F., Quiroz, J., and Glassman, E. (2017). *Clustering Latex Solutions to Machine Learning Assignments for Rapid Assessment.* Available online at: https://finale.seas.harvard.edu/files/finale/files/2017clustering_latex_solutions_to_machine_learning_assignments_for_rapid_assessment.pdf

Taoum, J., Nakhal, B., Bevacqua, E., and Querrec, R. (2016). "A design proposition for interactive virtual tutors in an informed environment," in *International Conference on Intelligent Virtual Agents* (Los Angeles, CA: Springer), 341–350. doi: 10.1007/978-3-319-47665-0_30

Tarcsay, B., Vasić, J., and Perez-Tellez, F. (2022). "Use of machine learning methods in the assessment of programming assignments," in *International Conference on Text, Speech, and Dialogue* (Brno: Springer), 341–350. doi: 10.1007/978-3-031-16270-1_13

Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., and Schmidt-Thieme, L. (2010). Recommender system for predicting student performance. *Proc. Comput. Sci.* 1, 2811–2819. doi: 10.1016/j.procs.2010.08.006

Tong, S., Liu, Q., Huang, W., Hunag, Z., Chen, E., Liu, C., et al. (2020). "Structure-based knowledge tracing: an influence propagation view," in *2020 IEEE International Conference on Data Mining (ICDM)* (Sorrento), 541–550. doi: 10.1109/ICDM50108.2020.00063

Toscher, A., and Jahrer, M. (2010). *Collaborative Filtering Applied to Educational Data Mining.* Washington, DC: KDD cup.

Trakunphutthirak, R., Cheung, Y., and Lee, V. C. (2019). "A study of educational data mining: evidence from a Thai university," in *Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33* (Honolulu, HI), 734–741. doi: 10.1609/aaai.v33i01.3301734

Tschuggnall, M., and Specht, G. (2013). "Detecting plagiarism in text documents through grammar-analysis of authors," in *BTW*, 241–259.

Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., and Ragos, O. (2020). Transfer learning from deep neural networks for predicting student performance. *Appl. Sci.* 10:2145. doi: 10.3390/app10062145

Ullah, F., Jabbar, S., and Mostarda, L. (2021). An intelligent decision support system for software plagiarism detection in academia. *Int. J. Intell. Syst.* 36, 2730–2752. doi: 10.1002/int.22399

UNESCO (2021). *When Schools Shut: Gendered Impact of COVID-19 School Closures.* UNESCO Publishing.

UNESCO (2022). *UNESCO's Education Response to COVID-19.*

United Nations (2015). *Goal 4: Quality Education.* United Nations.

Upadhyay, U., De, A., and Gomez Rodriguez, M. (2018). "Deep reinforcement learning of marked temporal point processes," in *Advances in Neural Information Processing Systems, Vol. 31* (Montreal).

Uto, M., and Okano, M. (2020). "Robust neural automated essay scoring using item response theory," in *International Conference on Artificial Intelligence in Education* (Ifrane: Springer), 549–561. doi: 10.1007/978-3-030-52237-7_44

Vani, K., and Gupta, D. (2014). "Using k-means cluster based techniques in external plagiarism detection," in *2014 International Conference on Contemporary Computing and Informatics (IC3I)* (Mysore), 1268–1273. doi: 10.1109/IC3I.2014.7019659

Varga, A., and Ha, L. A. (2010). "WLV: a question generation system for the QGSTEC 2010 task b," in *Proceedings of QG2010: The Third Workshop on Question Generation* (Pittsburgh, PA), 80–83.

Vijayalakshmi, V., Panimalar, K., and Janarthanan, S. (2020). Predicting the performance of instructors using machine learning algorithms. *High Technol. Lett.* 26, 49–54. doi: 10.5373/JARDCS/V12SP4/20201461

Villaverde, J. E., Godoy, D., and Amandi, A. (2006). Learning styles' recognition in e-learning environments with feed-forward neural networks. . *Comput. Assist. Learn.* 22, 197–206. doi: 10.1111/j.1365-2729.2006.00169.x

Vincent-Lancrin, S., and van der Vlies, R. (2020). *Trustworth Artificial Intelligence (AI) in Education: Promises and Challenges.* Organisation for Economic Cooperation and Development.

Vujošević-Janičić, M., Nikolić, M., Tošić, D., and Kuncak, V. (2013). Software verification and graph similarity for automated evaluation of students' assignments. *Inform. Softw. Technol.* 55, 1004–1016. doi: 10.1016/j.infsof.2012.12.005

Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S., and Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Comput. Hum. Behav.* 104, 106189. doi: 10.1016/j.chb.2019.106189

Walkington, C. A. (2013). Using adaptive learning technologies to personalize instruction to student interests: the impact of relevant contexts on performance and learning outcomes. *J. Educ. Psychol.* 105, 932. doi: 10.1037/a0031882

Wang, K., and Su, Z. (2015). "Automated geometry theorem proving for human-readable proofs," in *Twenty-Fourth International Joint Conference on Artificial Intelligence* (Buenos Aires).

Wang, L., Sy, A., Liu, L., and Piech, C. (2017). Learning to represent student knowledge on programming exercises using deep learning. *Int. Educ. Data Mining Soc.* doi: 10.1145/3051457.3053985

Wang, T., and Cheng, E. C. (2022). "Towards a tripartite research agenda: a scoping review of artificial intelligence in education research," in *Artificial Intelligence in Education: Emerging Technologies, Models and Applications* (Springer), 3–24. doi: 10.1007/978-981-16-7527-0_1

Wang, T., Su, X., Wang, Y., and Ma, P. (2007). Semantic similarity-based grading of student programs. *Inform. Softw. Technol.* 49, 99–107. doi: 10.1016/j.infsof.2006.03.001

Wang, Z., Lan, A., and Baraniuk, R. (2021). "Math word problem generation with mathematical consistency and problem context constraints," in *2021 Conference on Empirical Methods in Natural Language Processing.* doi: 10.18653/v1/2021.emnlp-main.484

Waters, A., and Miikkulainen, R. (2014). Grade: machine learning support for graduate admissions. *AI Mag.* 35, 64. doi: 10.1609/aimag.v35i1.2504

Wen, M., Yang, D., and Rose, C. (2014). "Sentiment analysis in MOOC discussion forums: what does it tell us?," in *Educational Data Mining 2014* (London).

Wester, E. R., Walsh, L. L., Arango-Caro, S., and Callis-Duehl, K. L. (2021). Student engagement declines in stem undergraduates during COVID-19-driven remote learning. *J. Microbiol. Biol. Educ.* 22, ev22i1-2385. doi: 10.1128/jmbe.v22i1.2385

Winthrop, R. (2018). *Leapfrogging Inequality: Remaking Education to Help Young People Thrive.* Brookings Institution Press.

Wolf, S., Aurino, E., Brown, A., Tsinigo, E., and Edro, R. M. (2022). *Remote Data-Collection During COVID 19: Thing of the Past or the Way of the Future.* World Bank Blogs. Available online at: https://blogs.worldbank.org/education/remote-data-collection-during-covid-19-thing-past-or-way-future

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., and Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *Int. J. Learn. Technol.* 4, 129–164. doi: 10.1504/IJLT.2009.028804

Woolf, B. P., Arroyo, I., Muldner, K., Burleson, W., Cooper, D. G., Dolan, R., et al. (2010). "The effect of motivational learning companions on low achieving students and students with disabilities," in *International Conference on Intelligent Tutoring Systems* (Pittsburgh, PA: Springer), 327–337. doi: 10.1007/978-3-642-13388-6_37

Wu, M., Mosse, M., Goodman, N., and Piech, C. (2019). "Zero shot learning for code education: rubric sampling with deep learning inference," in *Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33* (Honolulu, HI), 782–790. doi: 10.1609/aaai.v33i01.3301782

Wu, Q., Zhang, Q., and Huang, X. (2022). Automatic math word problem generation with topic-expression co-attention mechanism and reinforcement learning. *IEEE/ACM Trans. Audio Speech Lang. Process.* 30, 1061–1072. doi: 10.1109/TASLP.2022.3155284

Xie, X., Siau, K., and Nah, F. F.-H. (2020). COVID-19 pandemic-online education in the new normal and the next normal. *J. Inform. Technol. Case Appl. Res.* 22, 175–187. doi: 10.1080/15228053.2020.1824884

Xu, J., Han, Y., Marcu, D., and Van Der Schaar, M. (2017). "Progressive prediction of student performance in college programs," in *Thirty-First AAAI Conference on Artificial Intelligence* (San Francisco, CA).

Xue, K., Yaneva, V., Runyon, C., and Baldwin, P. (2020). "Predicting the difficulty and response time of multiple choice questions using transfer learning," in *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications* (Seattle, WA), 193–197. doi: 10.18653/v1/2020.bea-1.20

Yaneva, V., Baldwin, P., and Mee, J. (2020). "Predicting item survival for multiple choice questions in a high-stakes medical exam," in *Proceedings of The 12th Language Resources and Evaluation Conference* (Marseille), 6812–6818.

Yang, D., Sinha, T., Adamson, D., and Rosé, C. P. (2013). "Turn on, tune in, drop out: anticipating student dropouts in massive open online courses," in *Proceedings of the 2013 NIPS Data-Driven Education Workshop, Vol. 11* (Lake Tahoe, NV), 14.

Yang, Y., Shen, J., Qu, Y., Liu, Y., Wang, K., Zhu, Y., et al. (2020). "GIKT: a graph-based interaction model for knowledge tracing," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (Ghent: Springer), 299–315. doi: 10.1007/978-3-030-67658-2_18

Young, N., and Caballero, M. (2019). "Using machine learning to understand physics graduate school admissions," in *Proceedings of the Physics Education Research Conference (PERC)* (Provo, UT), 669–674.

Yudelson, M. V., Koedinger, K. R., and Gordon, G. J. (2013). "Individualized Bayesian knowledge tracing models," in *International Conference on Artificial Intelligence in Education* (Memphis, TN: Springer), 171–180. doi: 10.1007/978-3-642-39112-5_18

Yufeia, L., Salehb, S., Jiahuic, H., and Syed, S. M. (2020). Review of the application of artificial intelligence in education. *Integration* 12:1–5. doi: 10.53333/IJICC2013/12850

Zatarain-Cabada, R., Barrón-Estrada, M. L., Angulo, V. P., García, A. J., and García, C. A. R. (2010). "A learning social network with recognition of learning styles using neural networks," in *Mexican Conference on Pattern Recognition* (Puebla: Springer), 199–209. doi: 10.1007/978-3-642-15992-3_22

Zawacki-Richter, O., Marín, V. I., Bond, M., and Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education-where are the educators? *Int. J. Educ. Technol. High. Educ.* 16, 1–27. doi: 10.1186/s41239-019-0171-0

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., et al. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity* 2021, 8812542. doi: 10.1155/2021/8812542

Zhang, D., Wang, L., Zhang, L., Dai, B. T., and Shen, H. T. (2019). The gap of semantic parsing: a survey on automatic math word problem solvers. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 2287–2305. doi: 10.1109/TPAMI.2019.2914054

Zhang, H., Magooda, A., Litman, D., Correnti, R., Wang, E., Matsmura, L., et al. (2019). erevise: "Using natural language processing to provide formative feedback on text evidence usage in student writing," in *Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33* (Honolulu, HI), 9619–9625. doi: 10.1609/aaai.v33i01.33019619

Zhang, J., Shi, X., King, I., and Yeung, D.-Y. (2017). "Dynamic key-value memory networks for knowledge tracing," in *Proceedings of the 26th International Conference on World Wide Web* (Perth), 765–774. doi: 10.1145/3038912.3052580

Zhang, L., Zhao, Z., Ma, C., Shan, L., Sun, H., Jiang, L., et al. (2020). End-to-end automatic pronunciation error detection based on improved hybrid CTC/attention architecture. *Sensors* 20, 1809. doi: 10.3390/s20071809

Zhang, M., Baral, S., Heffernan, N., and Lan, A. (2022). "Automatic short math answer grading via in-context meta-learning," in *Proceedings of the International Conference on Educational Data Mining* (Durham).

Zhao, Y., Lackaye, B., Dy, J. G., and Brodley, C. E. (2020). "A quantitative machine learning approach to master students admission for professional institutions," in *International Educational Data Mining Society* (Virtual).

Zhao, Y., Ni, X., Ding, Y., and Ke, Q. (2018). "Paragraph-level neural question generation with maxout pointer and gated self-attention networks," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (Brussels), 3901–3910. doi: 10.18653/v1/D18-1424

Zhou, Q., and Huang, D. (2019). "Towards generating math word problems from equations and topics," in *Proceedings of the 12th International Conference on Natural Language Generation* (Tokyo), 494–503. doi: 10.18653/v1/W19-8661

Zhu, K., Li, L. D., and Li, M. (2021). A survey of computational intelligence in educational timetabling. *Int. J. Mach. Learn. Comput.* 11, 40–47. doi: 10.18178/ijmlc.2021.11.1.1012