The use of Artificial intelligence in school science: a systematic literature review

Heeg, Dagmar Mercedes; Avraamidou, Lucy

Published in:
Educational Media International

DOI:
10.1080/09523987.2023.2264990

Educational Media International

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/remi20

The use of Artificial intelligence in school science: a systematic literature review

Dagmar Mercedes Heeg & Lucy Avraamidou

To cite this article: Dagmar Mercedes Heeg & Lucy Avraamidou (2023) The use of Artificial intelligence in school science: a systematic literature review, Educational Media International, 60:2, 125-150, DOI: 10.1080/09523987.2023.2264990

To link to this article: https://doi.org/10.1080/09523987.2023.2264990

EDUCATIONAL MEDIA INTERNATIONAL
2023, VOL. 60, NO. 2, 125–150
https://doi.org/10.1080/09523987.2023.2264990

The use of Artificial intelligence in school science: a systematic literature review

Dagmar Mercedes Heeg and Lucy Avraamidou

Centre for Learning and Teaching, University of Groningen, Groningen, The Netherlands

ABSTRACT

Artificial Intelligence is widely used across contexts and for different purposes, including the field of education. However, a review of the literature showcases that while there exist various review studies on the use of AI in education, missing remains a review focusing on science education. To address this gap, we carried out a systematic literature review between 2010 and 2021, driven by three questions: a) What types of AI applications are used in school science? b) For what teaching content are AI applications in school science used? and, c) What is the impact of AI applications on teaching and learning of school science? The studies reviewed ($n = 22$) included nine different types of AI applications: automated assessment, automated feedback, learning analytics, adaptive learning systems, intelligent tutoring systems, multilabel text classification, chatbot, expert systems, and mind wandering detection. The majority of the AI applications are used in geoscience or physics and AI applications are used to support either knkowledge construction or skills development. In terms of the impact of AI applications, this is found across the following: learning achievement, argumentation skills, learning experience, and teaching. Missing remains an examination of learners' and teachers' experiences with the use of AI in school science, interdisciplinary approaches to AI implementation, as well as an examination of issues related to ethics and biases.

ARTICLE HISTORY
Received 29 May 2023
Accepted 17 September 2023

KEYWORDS
Artificial intelligence; systematic literature review; science education; learning; teaching

1. Introduction

The world has embraced Artificial Intelligence (AI) as a technology that can make our lives easier and can stimulate economic development (European Commission, 2020). As such, AI has been widely used in different domains of social systems. Education is no exception. Reports from UNESCO (Miao et al., 2021) and OECD (2021) provide evidence of the potential of AI to improve education by automating various aspects of teaching (e.g., administrative tasks, assessment, evaluation), personalizing learning (e.g., intelligent tutoring

CONTACT Dagmar Mercedes Heeg d.m.heeg@rug.nl Centre for Learning and Teaching, University of Groningen Nijenborgh 7, Groningen 9747AG, The Netherlands

systems), and making available new tools (e.g., virtual reality, augmented reality).

Over the last decade, researchers have examined the use of AI in teaching and learning for different purposes. For example, quite a few researchers examined the use of AI applications such as intelligent tutoring systems, virtual reality, and teaching evaluation (e.g., Chassignol et al., 2018; Gerard et al., 2015; Guan et al., 2020; Parsia et al., 2020; Zawacki-Richter et al., 2019; Zhai et al., 2020), as well as the use of more general pedagogical methods (e.g., game-based learning, collaborative learning) in teaching and learning in various contexts and disciplines (Bozkurt et al., 2021; Feng & Law, 2021; Zawacki-Richter et al., 2019).

However, a review of the literature showcases that while there exist various review studies on the use of AI in education, missing remains a review in the context of *science* education. For science education in particular, AI applications such as intelligent tutoring systems and automated assessment, may provide new ways to design engaging and personally meaningful learning environments which in turn could enhance students' interest in science. Specifically, school science (K-12) could benefit from more engaging, interactive, and inspiring learning environments, given students' declining interest in science (Van Griethuijsen et al., 2015). In order to examine the potential of AI applications in science education, this review study was guided by the following questions:

(1) What types of AI applications are used in school science?
(2) For what teaching content are AI applications in school science used?
(3) What is the impact of AI applications on teaching and learning of school science?

Through this systematic literature review, following the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (Moher et al., 2009), we aim to contribute to the existing knowledge base on the use of AI in science education by providing an in-depth overview of currently used AI applications for school science specifically and offer a set of recommendations for policy, research, and practice.

2. What is Artificial intelligence?

John McCarthy introduced the term Artificial Intelligence in the 1950s, defining it as "the science and engineering of making intelligent machines". Since then, the field has evolved and so has the definition of AI. Currently, a single universal definition of AI does not exist. This is partially due to the rapid development of the field, but also because researchers outside of computer science, in the areas of healthcare, philosophy, economics, and arts have taken a great interest in AI and its application. Each of these research fields has modified the term AI to its own needs and uses, and this interdisciplinary nature of AI makes it difficult to establish a consensus about what AI is.

In this review, we adhere to the more informal and broad definition of AI as proposed by Popenici and Kerr (2017), who define AI as "computing systems that are able to engage in human-like processes such as learning, adapting, synthesizing, self-correcting and use of data for complex processing tasks" (Popenici & Kerr, 2017, p. 2). We have chosen to use this definition because it fits the purpose of this review to gather insights into existing educational AI applications. We are less interested in AI technology itself, but rather in its applications, impact, and possibilities for school science. Hence, we prefer a broad definition of AI that applies to various intelligent applications and could include rather than exclude other applications.

To make this more concrete, we provide a description of two of the most common AI-enabled educational applications, and one none AI-enabled application. Intelligent Tutoring Systems (ITS) are an AI application as they provide personalized and adaptive instruction to learners. ITS take on the role of a human tutor, by offering guidance, feedback, and support to learners. It does so using AI techniques to capture learner's strengths, weaknesses, and progress. By adapting the learning content, difficulty level, and feedback, to these learning traits of each individual users, the learning experience is then optimized.

Similarly, automated assessment and feedback systems are considered AI as they autonomously assess students' work. One of the most important AI techniques for such systems is Natural Language Processing (NLP), which the applications use to analyze and understand human language. Using this technique, an automated assessment model can extract meaningful information from, for example, students' written responses to open answered text questions. Subsequently, this information can be assessed, using Machine Learning models that are trained to score and evaluate responses. These ML models are trained on large datasets of human-scored responses. What does not fall within AI applications, are pre-programmed educational robots. As a matter of fact, the majority of the currently used educational robots are pre-programmed, which means that they are designed to follow a fixed set of instructions or commands without any learning capabilities or adaptive behaviors. These robots often perform tasks such as following lines, light tracking, or dancing.

3. Methods

To develop an in-depth understanding of the currently existing AI applications and their impact on science teaching and learning, we conducted a systematic literature review, following the PRISMA guidelines (Moher et al., 2009). We used a set of inclusion criteria to identify relevant research articles (Xiao & Watson,

2019) in the following databases: ERIC, PsycInfo, Web of Science, and Scopus. We collected studies from the period of 2010 to 2021 as we aimed at gaining a more contemporary and timely understanding of the use of AI in science education.

3.1. Inclusion criteria

To be included in this review, a study must endorse the following criteria:

- The study must focus on an AI application
- The context of the study must be in science education
- The context of the study must be in primary or secondary education
- The study must perform primary or empirical research.

As a first step, we defined keywords that match the inclusion criteria, as for example, artificial intelligence, adaptive learning, intelligent tutoring system, science education, primary school, and secondary school. Following that, we expanded the set of keywords regarding the AI application with carefully chosen keywords from previously executed systematic literature reviews on AI applications in education (Chassignol et al., 2018; Gerard et al., 2015; Guan et al., 2020, Leo et al., 2020; Zawacki-Richter et al., 2019; Zhai et al., 2020). We followed this procedure to ensure the wide range of artificial intelligence applications in education (e.g., machine learning, expert systems, intelligent tutoring systems, automated scoring) is captured in the literature included. We then performed a search, reviewed the results, and adjusted the keywords. This process was repeated until we started receiving the same results. The final search string is presented in Table 1.

3.2. Databases and included studies

The search string was used to perform a search through four international databases: ERIC, PsycInfo, Web of Science, and Scopus. Given the scope and focus of these four databases, we argue that their combination provides an optimal search method that guarantees adequate coverage. The search covered the title, abstract, and keywords of every record, resulting in 539 initial hits. We observed that in these 539 initial hits, various educational journals were included with a focus on technology, such as *Computers and Education, International journal of Artificial Intelligence in Education, International, Journal of Research on Technology Education*, and *International Journal of Child-Computer Interaction*. However, very few science education journals were represented in the hits.

As this review paper aims to explore AI in the context of science education, and specifically at the school science level, we wanted to prevent the exclusion

Table 1. Search string used to screen four international databases for relevant literature.

Topic	Search string
AI application	"artificial intelligence" or "machine learning" or "neural network" or "adaptive teaching" or "adaptive learning" or "data mining" or "intelligent tutoring system" or "learning analytics" or "automated guidance" or "automated scoring" or "automated testing" or "expert system" or "intelligent retrieval" or "knowledge engineering" or "natural language processing" or "educational agent" or "automated tutor"
AND	
Science Education	"scientific argumentation" or "stem" or "science education" or "science curriculum" or "science teaching" or "science learning" or "physics" or "biology" or "chemistry" or "environmental science" or "ai curriculum" or "ai lesson"
AND	
Primary, secondary education	"primary school" or "secondary school" or "elementary school" or "middle school" or "primary education" or "secondary education" or "high school"

of relevant articles in science education journal by our specific search string, which is a common limitation of a systematic review methodology. We therefore decided to perform a very broad search, namely "Artificial Intelligence", on the individual websites of 10 established science educational journals: *Cultural Studies of Science Education*, *International Journal of Science Education*, *Journal of Computers in Mathematics and Science Teaching*, *Journal of Research in Science Teaching*, *Journal of Science Education and Technology*, *Journal of Science Teacher Education, Research in Science Education*, *Research in Science & Technological Education*, *Science Education, Science & Education*, and *School Science and Mathematics*. This resulted in 311 additional studies.

Thereafter, the 539 studies from the search in four international databases and the 311 studies from the Science Education journals were restricted to articles in the English language published no later than 2010 in highly ranked peer-reviewed journals. These restrictions guaranteed that the study focuses on the trends of AI in science education over the last decade. In order to enhance the quality of the studies reviewed we chose to include studies published only in peer-reviewed journals. However, we do acknowledge the limitation of not including studies published in peer-reviewed conferences.

Subsequently, duplicates were removed. This left 206 studies. We then screened the title and abstract of each study using the formulated inclusion criteria as a filter. Eventually, 50 studies were left. After that, the papers were fully read. 29 Studies did not adopt the inclusion criteria and were therefore excluded, resulting in 21 studies included for this review.

We excluded studies focusing on the use of educational robotics because as stated in the introduction the majority of such applications are not AI. However, there is a possible risk of missing out studies reporting on the use of more contemporary educational robotics that use AI.

For example, San Pedro, Baker, and Rodrigo's (2014) study, explored students' interaction with an intelligent tutoring system in the context of mathematics education instead of science education, and was therefore excluded. Ntemngwa and Oliver (2018) examined how middle school teachers integrated STEM lessons within courses that have a single subject science focus, using Robotics. Teachers made use of the Lego Mindstorms software and hardware to work on science and engineering problem solving using computer programming and robotics. The LEGO Mindstorm kits do not have built-in AI capabilities, nor was it explained in the study that teachers added this. Therefore, the study was not included in the review study as it did not focus on an AI application. Similarly, Ponticorvo, Rubinacci, Marocco, Truglio, and Miglino's (2020) study was excluded as they used the LEGO Mindstorm kits to explore how educational robotics can be used to foster and assess social relations in students' group.

Search alerts were created on ERIC, PsycInfo, Web of Science, and Scopus to inform us of any additional papers that would be published and fit the initial search string. On 15 April 2022, all search alerts had been gathered, fully screened, and accessed for eligibility, resulting in one additional study which was included in this systematic review. Leaving the total number of included studies in this review at 22 (Please see Figure 1 for a schematic overview of the search and inclusion and exclusion process).

3.3. Coding

To provide an in-depth overview of the AI applications used for school science in the identified 22 papers, the three proposed research questions, which focus on type of AI application, science teaching content, and impact on science teaching and learning, informed the analysis of the papers. First, all papers were fully read and summaries of each study were made that included the aim, type of AI application, method, findings, and conclusions. Subsequently, the summaries were used to determine for each individual paper the answers on the three proposed research questions.

To write a synthesis, all individual answers on the research questions were thematically analyzed, using a combination of inductive and deductive coding techniques. An example of the coding process is presented in Table 2. Themes emerged within the codes, and were used to write the synthesis offered in the Result section.

4. Results

4.1. RQ1: what types of AI applications are being used in school science?

The studies reviewed included nine different types of AI applications: *automated assessment, automated feedback, learning analytics, adaptive learning systems, intelligent tutoring systems, multilabel text classification, chatbot, expert system*, and *mind wandering detection*. These applications range from simulating conversations with human users (i.e., chatbot) to the use of attention control mechanisms (i.e., mind wandering detection). The studies typically used to

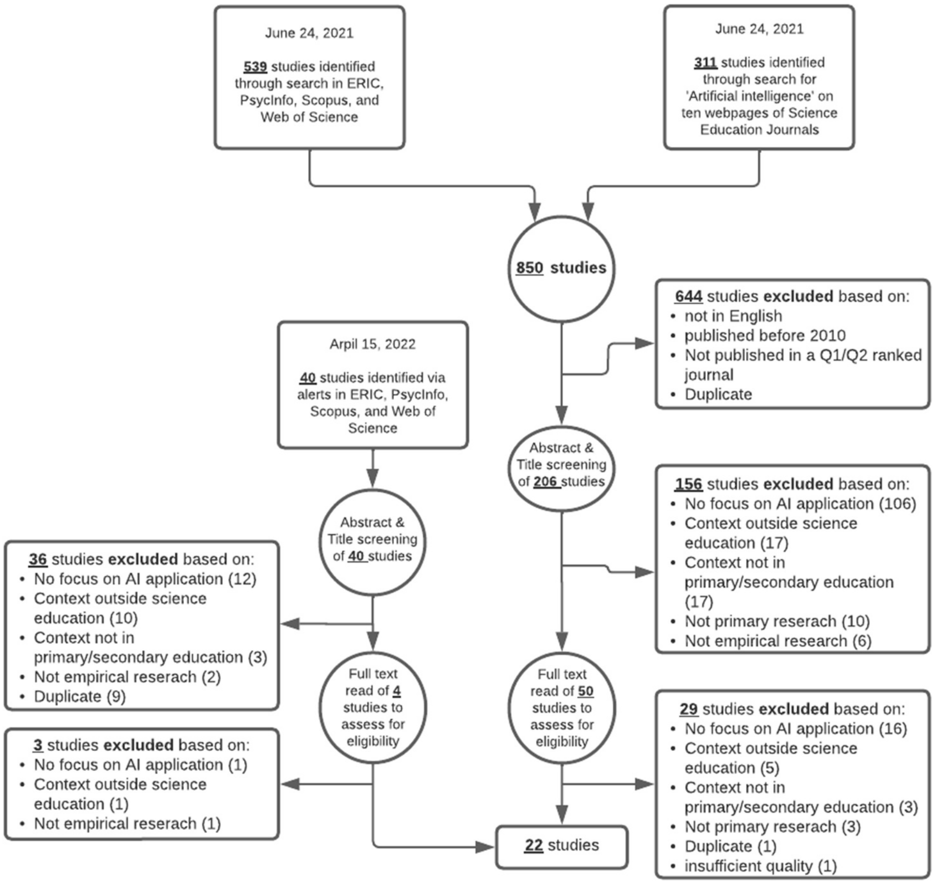

Figure 1. Prisma diagram (inspired by Moher et al., 2009; Zawacki-Richter et al., 2019).

examine the ways in which AI tools are used and not the impact that these tools might have on learning. In reviewing these studies, we observed that they often lacked an explicitly articulated definition of the AI application. Therefore, instead of definitions provided by the included studies, we provided self-formulated descriptions of each application in Table 3. In five cases, two types (i.e., automated assessment and automated feedback) of AI applications were covered in the paper. Table 3 included these studies under both AI types.

In the reviewed literature, automated assessment and feedback applications were well represented, together with personalized learning systems and predictive modeling. This is in line with previous identified AIEd research trends over the last decade (Bozkurt et al., 2021; Feng & Law, 2021; Zawacki-Richter et al., 2019). Other applications for school science education that are less represented in the literature are chatbots, multilabel text classifiers, mind wandering detection systems, and expert systems. These varied applications found in the reviewed literature are in line with the wide range of AI in general.

Table 2. Sample of coding process. A sample of the data assigned to a particular research question is coded and the theme(s) that emerged from these codes are given.

Research Question	Sample data	Codes	Theme(s)
1. AI application	**Automated scoring** and **feedback** are provided when students construct responses to explanation and uncertainty attribution prompts for each argumentation block. Students' responses are scored immediately following submission to c-rater-ML (Zhu et al., 2017, p. 1654)	Automated scoring Automated feedback	Automated assessment Automated feedback
	To identify how students' responses changed after **receiving immediate feedback** statements, we focused on variables related to argumentation revisions. (Zhu et al., 2017, p. 1659)	Student receiving feedback	Student support
2.Teaching content	Students completed one of **four FOSS modules** (*Variables, Magnetism,* **and** *Electricity, Measurement, and Water*) and were tested pre-post with the FOSS-ASK assessment for that module (Ward et al., 2013, p. 1120)	Existing physic modules	Fostering domain-specific knowledge Existing learning materials
	HASbot was developed to support students' revision of **scientific arguments** in real-time when they were engaged in classroom activities (Lee et al., 2019, 616)	Scientific argumentation	Fostering scientific skills
3. Impact	The experimental results showed that the **students' learning achievements were significantly improved** in terms of several cognitive processes in Bloom's taxonomy of educational objectives with the assistance of this real-world learning guidance approach (Wu et al., 2013, p. 227)	Learning achievements	Students' learning achievements
	When the students' **opinions** about the chatbot were analyzed, it could be summarized that they generally found **this activity** useful and **enjoyable**; it helped them learn, it was **fun**, it increased their curiosity about the science course, they could reach it whenever they wanted, and they wanted to use it in other courses such as English, mathematics and social sciences (Topal et al., 2021, p.20)	Opinion	Students' experiences

Table 3. Overview and examples of different types of AI applications covered in the 22 papers. When a paper covered multiple AI applications, they are included under multiple AI applications.

Type of AI application	Description	offers support to students	offers support to teachers	# of papers that covered the ai application
automated assessment	AI algorithms that can automatically assess written text.	x	x	7
automated feedback	AI algorithms that can automatically give feedback on input.	x	x	5
Intelligent tutoring system	A system that aims to improve students' learning achievements by providing immediate tailored feedback to students.	x		3
Predictive modeling	Models that gain more insights on student learning by analyzing data about students and their learning process (e.g., achievement, behavior).		x	3
adaptive learning system	A system that aims to improve students' learning achievement by tailoring the learning content to individual students' learning, cognitive abilities, and achievements.	x		3
mind wandering detection	A system that detects when students' attention shifts from task-related to task-unrelated thoughts, also known as "mind wandering" (Hutt et al., 2019)		x	2
expert system	"A system that is an artificial intelligence program designed to simulate expert reasoning based on the knowledge elicited from domain experts" (Chu et al., 2010, p. 280).	x		2
multi label text classification	Applications that can assign multiple labels to textual data.		x	1
chatbot	A computer program that is designed to interact with users by simulating a conversation through written text.	x		1

The majority of the AI applications were designed to offer support to students, meaning that students directly interact with the applications, receiving, for example, adapted learning content, automated feedback, or answers to their questions (i.e., chatbot). AI applications that offer support to teachers often aimed to save teachers time by automating tasks such as assessment, or to help teachers plan teaching interventions for struggling students through predictive modeling applications.

4.2. RQ2: for what teaching content are AI applications in school science used?

Table 4 provides a detailed overview of the teaching content in which the AI applications are used. The study by Yağci and Çevik (2019) is excluded from this table, as their AI application predicts students' achievements in a full science program and is not used in a specific course. From the other 21 studies, the majority of the AI applications are used in in the context of geoscience ($n = 5$) or physics ($n = 9$) lessons. Despite a strong influence/authority of these two domains, the teaching content of the AI applications reviewed is diverse, as is evidenced by the large variety in subtopics of the physics and geoscience learning activities in Table 4. Less frequent teaching content in the reviewed literature are science, biology, and AI.

Regarding the learning objectives accompanying the teaching content, the AI applications are used to either teach domain-specific knowledge to students or to develop scientific argumentation skills. For example, the automated assessment and feedback application by Zhu et al. (2017), and further developed versions of the application by Mao et al. (2018) and Lee et al. (2019, 2021), aimed to improve the scientific argumentation skills of students. These applications provided a curriculum during which students were asked to give multiple scientific argumentations. After submitting an argument, students received automated assessment and feedback on how to improve these arguments. Subsequently, the students were provided with an opportunity for revisions on which they would again receive a score and feedback. Overall, the automatic assessment and feedback system trained and improved the argumentation skills of students.

The intelligent tutoring system by Dolenc et al. (2015), illustrates applications focusing on knowledge transfer. This tutoring system taught the Gears subject matter, which serves as the last chapter of the Technology and Science Course in the eighth grade of an elementary school in Slovenia. By providing students with the content of this subject matter, and proposing individually adapted learning paths based on students' cognitive abilities, students were able to master the content of this Technology and Science Course chapter.

In the reviewed literature, the learning materials of AI applications were often derived from already existing learning materials based on the national curricula.

Table 4. An overview of the science teaching content AI applications are embedded in.

Authors (year)	AI application	Course/educational level	Module/learning activity	Objectives
An overview of the science teaching content AI applications are embedded in or used for				
Huang et al. (2011)	Automated assessment & Feedback	Science/Primary school	The AI application was embedded into a Natural science course	To improve students' argumentation skills
Zhu et al. (2017)	Automated assessment & Feedback	Geoscience/Grade 9–12	The AI application was embedded into a modified version of the High-Adventure Science "What is the future of Earth's Climate?" module. Eight scientific argumentation tasks embedded in five different modules: 1) Earth's changing climates, 2) Solar radiation, 3) relation between levels of atmospheric carbon dioxide and water vapour, 4) ice and clouds and Earth's temperature, 5) global warming.	To improve students' argumentation skills
Mao et al. (2018)	Automated assessment & Feedback	Geoscience/ Secondary school	The AI application was embedded into an Earth science module in which students explore factors related to climate change	To improve students' argumentation skills
Lee et al. (2019)	Automated assessment & Feedback	Geoscience/Grade 8–12	The AI application was embedded in a modified version of the High Adventure Science "Will there be enough fresh water?" module. Eight scientific argumentation tasks embedded in six different modules: 1) constructing an agument, 2) availability of freshwater, 3) using freshwater, 4) groundwater movement, 5) groundwater and surface water, 6) using groundwater wisely.	To improve students' argumentation skills
Lee et al. (2021)	Automated assessment & Feedback	Geoscience/Grade 8–9	The AI application was embedded in a modified version of the High Adventure Science "Will there be enough fresh water?" module. Three specific scientific argumentation tasks (Trap, Aquifer, Supply) embedded in six different modules: 1) constructing an agument, 2) availability of freshwater, 3) using freshwater, 4) groundwater movement, 5) groundwater and surface water, 6) using groundwater wisely.	To improve students' argumentation skills
Sung et al. (2021)	Automated assessment	Physics/Grade 9	The AI application was embedded into the "Two Thumbs Up" module, that focuses on heat conduction and thermal perception of heat transfer.	To help assess students' content knowledge of the module
Maestrales et al. (2021)	Automated assessment	Physics & Chemistry/ Secondary school	The automated assessment model was trained on data (students' test answers) of four question from the National Assessment of Education Progress (NAEP) open-source science test-bank items. The questions were related to physics or chemistry everyday scenarios.	To assess text written answers by students

(*Continued*)

Table 4. (Continued).

Authors (year)	AI application	Course/educational level	Module/learning activity	Objectives
Dolenc et al. (2015)	Intelligent Tutoring System	Physics/Grade 8	The ITS is called TECH8 and included the Gears subject matter, which is the last chapter of a Technology and Science course in Slovenia.	To support students in learning the content of the Gears subject matter.
Ward et al. (2013)	Intelligent Tutoring System	Physics/Primary school	The ITS was used as a supplement to normal classroom instruction using the Full Option Science System (FOSS). There are 26 FOSS modules and are used in every state in the United States. The ITS was used on four modules: Variables, Magnetism, Electricity, Measurement & Water.	To support students in learning the content of the four FOSS modules.
Zulfiani et al (2018, 2021).	Adaptive learning system	Physics/Secondary school	The teaching materials taught with the adaptive learning system are the materials on movement and simple machines Indonesia's 2013 National Curriculum.	To support students in learning the content of the movement and simple machine content.
Mead et al. (2019)	Adaptive learning system	Geoscience/ Secondary school	The adaptive learning system in embedded in an immersive, interactive virtual field trip (iVFT) on the Nilpena. The lesson centers on a series of fossil beds from the Nilpena site in the Flinders Ranges of South Australia.	To support students' learning during the Nilpena iVFT.
Gaudioso et al. (2012)	Predictive modeling	Physics/Secondary school	The predictive model is embedded into an adaptive learning system for teaching Physics.	To support teachers in planning learning interventions for students.
Wen et al. (2018)	Predictive modeling	Physics/Grade 10	The predictive model is embedded into a problem-solving game on Physics. The game asks students to model how the velocities of two objects on a cliff will change after collision.	To support teachers by providing detailed insights in students' learning processes
Udandarao et al. (2021)	Multi Label Text Classification	Physics/Grade 6–12	This text classification system was used on a physics course materials from the federally supported CBSE curriculum in India.	To label text from a physics books (e.g., definition, example, effect), to sort textbook text.
Topal et al. (2021)	Chatbot	Science/Grade 5	The chatbot was embedded in the "Matter and changing state of matter" unit of a Science course.	To support students' learning on the topic of Matter.
Wu et al. (2013)	Expert System	Geoscience/ Secondary school	The expert system is embedded into an existing learning activity of the sample school. The learning activity was embedded into the Geoscience course and aimed to train students to identify and differentiate the features of a set of target rocks.	To support students' skills of identifying a set of rocks.
Hutt et al. (2019) & Bosch, D'Mello (2021)	Mind wandering detection system	Biology/Secondary school	The mind wandering detection system is embedded into a regular biology class that includes the intelligent tutoring system GuruTutor that focuses on introductory biology topics (e.g., osmosis, protein function).	To support students' learning during GutuTutor.
Gonzalez et al. (2017)	Intelligent tutoring system	AI education/ Primary and secondary school	The ITS is embedded in an AI exhibit. The ITS enables an intelligent guide to engage with museum visitors and teach them about the Turing Test. (note: the researchers include this under informal science education).	To engage visitors with the exhibit and support their learning on the Turing test.
Nyeet al. (2021)	Expert system	Science/Secondary school	The expert system is embedded in MentorPal, a virtual agent that offers informational interviews with professionals at, for example, career fairs.	To increase students' understanding about possible careers.

Hence, AI is being used to teach content that was already being taught. AI applications are often merely used to digitalize or automate existing learning materials, instead of searching for a new and innovative way of teaching and learning that utilizes the opportunities AI brings to the educational field.

4.3. RQ3: what is the impact of AI applications on teaching and learning school science?

In this section, we synthesize the impact of the AI applications school science as reported by the reviewed literature. Common evaluation strategies to measure the impact of the AI applications were a control group and experimental group, pre-test and post-test, log data analysis, or human-machine agreement analysis. In a few cases ($n = 6$), the impact was additionally examined with qualitative data obtained from surveys, interviews, video recordings, or audio recordings.

The articles are, based on the nature of their impact, subcategorized into four themes: impact on i) students' learning achievements; ii) students' scientific argumentation skills; iii) students' learning experiences; and, iv) teaching. These four subsections elaborate more on the subcategorized type of impact and present tables in which all related studies are summarized.

4.3.1. Impact on students' learning achievements

Seven studies reported an impact on students' learning achievements (Table 5). Overall, the impact of AI applications integrated into science curricula was satisfactory on students' learning achievements, meaning that students achieved the learning objectives of the curriculum in which the AI applications were integrated. In a few cases ($n = 4$), authors argued that the educational approaches using AI applications could impact students' learning achievements as good as, or better than, conventional teaching approaches. No reports of negative influences on learning achievements were discussed or observed in the reviewed literature.

4.3.2. Impact on students' scientific argumentation skills

Four articles reported a positive impact on students' scientific argumentation skills through the use of an automated assessment and feedback system. The relevant articles are presented in Table 6. All articles either used a pre-test/post-test comparison, or a log data analysis to illustrate the impact on students' scientific argumentation skills. In one case, it was argued that a scientific argumentation curriculum with AI features facilitated better learning improvements than the same curriculum without AI features. Again, no negative impact or lower impact compared to other teaching methods were reported.

Table 5. Summaries of studies that reported impact on students' learning achievements.

Authors (year)	Type of AI application	Purpose	Method	Findings regarding impact on students' learning achievements
What is the impact of AI applications on students' learning achievements?				
Dolenc et al. (2015)	Intelligent Tutoring System	To evaluate and optimize Intelligent Tutoring System called TECH8	Quantitative method; post summative-assessment knowledge test; $n = 78$	The use of TECH8 resulted for the majority of the students in higher learning results. However, there were a few students who achieved lower results while using TECH8 than students who received conventional teaching
Ward et al. (2013)	Intelligent Tutoring System	To design and evaluate an Intelligent Tutoring System called Marni, that can help support students in developing skills to reason and talk about science.	Quantitative method; Quasi-experimental design; pre- and post-test; three-point rating scale survey on experience and impressions $n = 1478$	The Intelligent Tutoring System had roughly the same effect on students' learning outcomes as a human tutor has.
Wu et al. (2013)	Expert system	To develop an expert system to support context-aware ubiquitous learning activities that engages students in higher order thinking ability.	Quantitative method; experimental study; pre- and post-test; $n = 58$	Students using the expert system (experimental group) had significantly higher learning achievements than the control group who did not use the expert system.
Zulfiani et al. (2018)	Adaptive learning system	To develop a learning system that adapts to different learning styles (visual, aural, reading/writing, kinesthetic) to improve students' learning.	Mixed-method; pre-interviews and questionnaires, post-test score; $n = 30$	The adaptive learning system was only determined to effectively teach the physics concepts, when it adapted to a kinesthetic learning style.
Zulfiani et al. (2021)	Adaptive learning system	To develop an Android version (available on smartphones) of the adaptive learning system designed by Zulfiani et al. (2018).	Quantitative method; quasi-experimental; pre- and post-test; $n = 114$	A significant difference in learning outcomes between the Android version of the adaptive learning system, available on smartphones, and the original version, only available on laptops or PCs.
Mead et al. (2019)	Adaptive learning system	To validate the effect of a virtual field trip that includes an adaptive learning system on students' learning achievements.	Mixed-method; pre-, post-, and retaining test, open-ended question survey; $n = 153$	After participating in the virtual field trip, there was a significant difference between pre- and post-test, and between pre- and retaining test.
Topal et al. (2021)	Chatbot	To investigate the effect of chatbots on students' success and students' opinions about chatbots.	Mixed method; pre-test and post-test, semi-structured interviews; $n = 41$	No significant difference in academic achievement between the experimental group that had access to the chatbot and control group. Hence, the chatbot does not offer any learning gains compared to not having a chatbot.

Table 6. Summaries of all studies that reported impact on students' scientific argumentation skills.

Authors (year)	Type of AI application	Purpose	Method	Findings regarding impact on students' argumentation skills
What is the impact of AI applications on students' scientific argumentation skills?				
Huang et al. (2011)	Automated assessment & feedback	To design an automated assessment and feedback system that teachers can use to assess students' scientific argumentation and lighten teachers' workload.	Quantitative method; experimental study; pre- and post-test; $n = 31$	Significant difference between the argumentation level before and after the curriculum in the experimental group and a significant difference between the experimental and control group. Hence, the tool is an effective learning tool to develop students' scientific argumentation.
Zhu et al. (2017)	Automated assessment & feedback	To investigate the impact of automated feedback on student learning of scientific argumentation	Quantitative method; log data (e.g., # of revisions, initial task scores, end task scores); $n = 183$	Results showed that students who revised their scientific arguments based on the given feedback had significantly higher final scores than students who did not revise their arguments.
Lee et al. (2019)	Automated assessment & feedback	To support students' in developing their uncertainty-infused scientific argumentation skills	Mixed method; log data (e.g., # of revisions, initial scores, final scores), pre- and post-test, video recordings; $n = 300$	Significant improvement in students' uncertainty-infused scientific argumentation between pre- and post-test.
Lee et al. (2021)	Automated assessment & feedback	To examine whether simulation feedback supports students in developing their scientific argumentation skills	Quantitative method; log data (e.g., # of revisions, initial and final scores),; $n = 343$	Students who received simulation feedback were more likely to rerun simulations and improve their scientific arguments.

4.3.3. Impact on students' learning experiences

It is worth noting that only six studies provided empirical evidence related to the students' learning experiences while working with the AI applications. Although almost all studies strive to evaluate their AI application, this evaluation is often focused on accuracy or effectiveness, and little attention is given to the impact on the learning experiences of students.

The six studies that examined students' learning experiences used surveys, interviews, or video data, to gather a broad idea of how students themselves perceived their experience in terms of engagement, enjoyment, and motivation. However, no validated tools were used to measure learning-related constructs such as motivation, interest, and attitudes. Moreover, the studies did not use experimental research designs to distinguish effects on learning experiences by the actual AI features from the effects of the curriculum in which the AI features were generated, hence it is not clear if the impact on students' learning experiences can be attributed to the AI features or from other features of the curriculum students participated in.

In Table 7, we provide an overview of the impact of the AI applications on students' learning experiences as reported by the reviewed literature because these initial, limited findings may propose some useful insights for future research.

4.3.4. Impact on teaching

In this subsection, we present the findings of seven studies that described a potential impact of AI applications on teaching (Table 8). An important caveat is that none of the seven studies examined the impact on teaching in particular or teachers' practices. Instead, the studies provide evidence for the potential impact and discuss how teachers, in theory, could use this in the classroom. Hence, the research so far is limited to design studies, where the focus lays on developing and validating AI applications. Although researchers seem optimistic about the impact on teaching, there is no concrete evidence of the impact of AI on the nature of science teaching.

Overall, the studies reported three different ways to impact teaching. Firstly, AI applications can support teachers by taking over time-consuming tasks. This can save time for teachers, which can relieve work pressure on teachers, or give them more time to spend on individual teaching. Secondly, automatic assessment based on machine learning techniques can help teachers assess students on skills that were previously too time-consuming to assess. And finally, the third way to impact teaching is via predictive modeling which can support teachers in planning teaching intervention for more optimal impact.

Table 7. Summaries of all studies that reported impact on students' learning experiences.

Authors (year)	Type of AI application	Purpose	Method	Findings regarding students' learning experiences
What is the impact of AI applications on students' learning experiences?				
Ward et al. (2013)	Intelligent Tutoring System	To design and evaluate an Intelligent Tutoring System called Marni, that can help support students in developing skills to reason and talk about science.	Quantitative method; Quasi-experimental design; pre- and post-test; three-point rating scale survey on experience and impressions $n = 1478$	Marni was perceived as a positive learning experience by a majority of the students. Most students enjoyed working with Marni and reported an increased science interest and thought Marni helped them with learning science.
Mead et al. (2019)	Adaptive learning system	To demonstrate the effectiveness of an immersive, interactive, virtual field trip, in achieving learning objectives	Mixed-method; pre-, post-, and retaining test, open-ended question survey; $n =$ 153	Most students reported to have felt engaged and interested with the adaptive learning system and a third reported to have enjoyed learning with it.
Topal et al. (2021)	Chatbot	To investigate the effect of chatbots on students' success and students' opinions about chatbots	Mixed method; pre-test and post-test, semi-structured interviews; $n = 41$	The chatbot was perceived as a fun experience, and increasedstudents' curiosity for learning science. Students thought the chatbot could be a useful additional learning tool both in and outside the classroom, but did not see it as a competent replacement for the teacher.
Gonzalez et al. (2017)	AI exhibit	To develop an engaging museum exhibit supporting informal science education	Qualitative method; interview, surveys, and observation; $n = 38$	The time duration of the exhibit was concluded to be too long, but overall the exhibit was determined to be able to engage visitors and offer them an enjoyable experience.
Hutt et al. (2019)	Mind wandering detection system	To develop attention-aware learning technologies as a mean to improve students' attentional focus, engagement, and learning.	Quantitative method; mind wandering warnings by system, self-reporting of mind-wandering by students; $n = 38$	The accuracy of mind-wandering detection by the system was not significantly accurate yet. Therefore, no impact on students' learning experience in the classroom was reported.
Lee et al. (2019)	Automated assessment & feedback	To support students' in developing their uncertainty-infused scientific argumentation skills	Mixed method; log data (e.g., # of revisions, initial scores, final scores), pre- and post-test, video recordings; $n = 300$	The AI application helped students identifying which elements were missing from their scientific arguments. Moreover, the design of the rainbow-colored bars, motivated students to continue revising their answers until there score was in the green (success) area of the bar.

Table 8. Summaries of all studies that reported impact on teaching.

Authors (year)	Type of AI application	Purpose	Method	Findings	Potential impact on teaching as argued by authors
AI applications and their reported Impact on teaching					
Udandaraoet al. (2021)	Multi-label text classification	To create a multi-label text classification system, that can be used to build an automated question-answering system for educational content.	Quantitative method; comparison study between the labels assigned by the multilabel text classification, seven other machine learning techniques, and human coders	The developed multi-label text classification system outperformed all other machine and deep learning techniques, and had a satisfying accuracy level when compared with the human coders. Hence, the labeling system functions well.	In the future, this type of classification system might be used to build a question-answering system. This will save teachers a lot of time, as they don't have to develop exam questions anymore.
Sunget al. (2021)	Automated assessment	To investigate how machine learning could improve the automated assessment of multi model representational thinking.	Quantitative method; comparison of human and machine assigned scores;	The developed model to assess multi model representational thinking outperformed a more traditional machine learning model, and had a satisfying accuracy compared to human coders.	The proposed automated assessment model can save teachers time, by taking over assessment tasks. Moreover, assessing multi model representational thinking is normally too time consuming to teach and assess by teachers. Hence, this automated assessment tool makes it easier to teach these specific skills.
Maestraleset al. (2021)	Automated assessment	To validate a new automated assessment model to assess constructed response items as an alternative to human.	Quantitative method; comparison human and machine assigned scores	There was a high human-machine agreement in analyzing students' constructed responses. Meaning that this tool can be used to automate assessment of constructed responses.	The automated assessment application can save time for teachers by taking over assessment tasks.
Yagci and Çevik (2019)	Predictive modeling	To predict the academic achievements of students in science courses and to put forth the measures to be taken against their failure.	Quantitative method; questionnaires and students' average science scores; n=1972	The predictive model was able to accurately predict students' academic achievements with an accuracy between 95 and 98 percent. Moreover, the model identified the most important factors affecting the academic achievements of students (e.g., no space at home for a study or a private bookshelf, crowded classes)	May help teachers improve the learning process of students, by helping them identify which students need intervention and when and which problems need to be addressed.
Gaudioso et al.(2012)	Predictive modeling	To develop a tool that assists teachers in adaptive educational systems and helps them identify students that need intervention and show the usefulness of such tool.	Quantitative method; three data sets including, among other things, final grades and prior knowledge	The predictive model was too simple to provide useful and validated information on students' learning behavior. The model was able to find a few signals for students failing, such as spending less than ten minutes studying in the online learning environment and receiving a low initial test score.	Tools to identify students at risk of failing may support teachers in deciding which students need extra teaching intervention.

(*Continued*)

Table 8. (Continued).

Authors (year)	Type of AI application	Purpose	Method	Findings	Potential impact on teaching as argued by authors
Wen et al. (2018)	Predictive modeling	To develop a model that analyzes students' activity patterns while working on a scientific problem to better understand students' modeling progress and students' strategies for solving science problems.	Quantitative method; data from successful and unsuccessful students	The model was able to identify certain learning behaviors from successful and unsuccessful students.	Teachers may use this model to gain insights on students' learning behaviors and to adjust their teaching or plan useful interventions for students, when they show learning activity associated with unsuccessful results.
Nye et al.(2021)	Expert system	To examine a virtual mentor system is a feasible replacement of real life mentors	Mixed method; log data, postonly survey on usability, preand post-test survey on students' career knowledge and attitudes; n=31	MentorPal was perceived as usable tool for five to twenty minutes of question and answering. Students perceived their career knowledge to be improved.	With virtual mentors, teachers do not longer have to spent a lot of time on finding volunteers to share experiences. Organizing career fairs becomes easier and more feasible to offer to students.

5. Discussion

Based on the review of the 22 selected studies, our first conclusion is that the literature on AI applications in school science follows similar trends as the review studies on AI application in education, in general. Similar to the corpus of Zawacki-Richter et al. (2019), who reviewed 146 studies on AI applications in higher education, the majority of the studies took placed in the United States and Asia. Moreover, the three most dominant themes of AI applications identified in our review are: automated assessment and feedback, personalized learning, and predictive modeling. These are well aligned with previously identified themes and trends in research on AI in education (Bozkurt et al., 2021; Feng & Law, 2021; Guan et al., 2020). The similarities between the research themes identified in broad review studies on AI in education and the themes in this more specific systematic review focused on AI in school science education, possibly imply that researchers have a similar approach to integrating AI in school science education as in other educational domains.

A second conclusion we draw from the corpus analysis is that the impact of AI applications on school science shows great potential in tackling three major challenges in science education. First, AI applications can help lighten the heavy workload of science educators by taking out time-consuming tasks such as assessment and feedback. Secondly, AI applications can enhance students' general low interest in science by providing them with engaging and entertaining intelligent tutors, or non-traditional personalized learning environments. Finally, AI applications can help optimize teaching processes to improve the general low learning achievements in science subjects. Applying a more critical lens towards these findings, showcases that AI applications have been used either for more practical reasons (i.e., reducing workload, automatizing feedback) or for supporting students' interest and conceptual understanding. Missing remains engagement with goals related to how students interact with AI applications, how AI applications might transform learning environments to promote inclusivity goals, as well as how AI tools might be used to enhance socio-emotional learning.

A third conclusion that can be drawn from our synthesis of the findings of the reviewed studies is that the overall positively reported impact of AI application on school science education lacks concrete empirical evidence. In the reviewed literature AI applications are declared as successful or the impact of the application is determined to be positive when the application works sufficiently. However, in the reviewed literature "working sufficiently" was narrowed down to a certain level of accuracy, or certain values of the post-test. A more holistic approach to learning experiences which would enable a more in-depth understanding of how AI applications might be used to transform school science is missing in the literature. For example, the studies on predictive models reported positive outcomes since the predictions of the models were accurate. However, the researchers tended to assume that these accurate prediction models will automatically have a positive impact on student learning and achievement. Moreover, missing remains an understanding of *how* AI tools might support science engagement and learning of *all* learners, which directly connects to sustainability goals related to inclusivity

This brings to light a first major limitation of the reviewed corpus: a lack of qualitative empirical studies that would provide insights into the ways in which AI tools might support science teaching and learning as well as multiple types of interactions that take place in the classroom: students with AI applications, students with other students, students with teachers, and teachers with AI applications. Instead, the majority of the studies reviewed have used quantitative methods that lack depth and attention at the microlevel aspects of learning (e.g., discourse analysis, interaction). Moreover, students' and teachers' experiences have not been examined thoroughly. Few studies used a mixed-method approach, where little additional qualitative data were collected, often from a very short period and/or limited sample size. This lack of qualitative and longitudinal studies was also observed in the literature on AI applications in higher education by Zawacki-Richter et al. (2019).

A second limitation that we identified in reviewing the existing knowledge base on the use of AI in school science is attention to issues related to bias and ethics, especially in relation to race and gender. This is a striking finding considering the increasing attention on trustworthy and ethical AI from researchers and policymakers globally. Ethical considerations should not only be taken into account during the implementation of AI tools but also during the development and implementation to ensure safe and fair AI applications for school science.

Finally, the last conclusion that we draw from our corpus analysis is that in the past 11 years, AI applications have been used in school science to automate or optimize already existing educational practices. Clear examples of this automatization are AI applications such as automated assessment and feedback, which are specifically designed to take over tasks from teachers. Examples of optimization include predictive models, which provide teachers with information on students' behavior or achievements to schedule an effective intervention. Other less obvious examples of automatization and optimization are AI applications such as intelligent tutoring systems, or adaptive learning systems. These personalized learning tools try to optimize students' learning process and help them achieve the learning objectives of the curriculum. The teaching content that these applications are used for is almost exclusively existing science curricula. Teaching content from textbooks is transferred to an online learning environment with some personalized learning features. In all instances, the essence of the teaching content and the learning objectives remained the same, only the delivery of the teaching content to the students changed. Thus, AI applications are being integrated into the existing science curriculum without any change in the learning objectives, which is both fragmented and limited

We argue that a sole focus on optimizing and automating educational processes is a third limitation of the corpus. Missing remains an understanding of how teachers and AI tools might work in synergy to support science engagement and learning. We recommend that researchers, policy-makers, and educators should aim for a new human-machine alliance, where teachers' abilities to recognize emotions, communicate, and offer mental support to students are combined with AI features focusing on analyzing student behavior, assessing, evaluating, etc. To do this, interdisciplinary approaches to research are needed to investigate how AI can truly support and extend the way teachers teach and redefine school science education. Or, as Miao, Holmes, Huang, and Zhang argued (2021), we should investigate how we can use "the unique affordances of AI to reimagine teaching and learning" (p. 19).

6. Implications

In the introduction, we identified AI as a potential technology that can improve science education. Though we cannot predict exactly the future types and uses of AI applications in science education, it is clear that AI has already transformed teaching and learning. The synthesis of the literature holds important implications for future research on AI in science education, both for theory and practice. These implications also point to the possible future use of AI in science education and the potential risks associated with it especially in relation to fairness and ethics.

6.1. Examining students' and teachers' experiences through explorative qualitative studies

The large majority of the literature studies are quantitative of nature and focused on validating the AI application in terms of accuracy or efficiency. For AI in school science to move forward, we need more explorative qualitative studies on students' and teachers' experiences with AI applications in the classroom, in order to get a better understanding of *how* AI applications might support science engagement and learning.

6.2. Addressing the ethical risks on AI in science education through guidelines and AI education

Despite the growing attention to the ethical challenges related with artificial intelligence, none of the reviewed studies discussed ethical issues that were considered during the design or implementation phase. To ensure safe, fair, and impartial AI applications in school science we urge AI science curriculum designers to take into account the seven key requirements that AI system should meet to be trustworthy, according to the report "the Ethics guidelines of Trustworthy AI" developed by the High-Level Expert Group on Artificial Intelligence, and that was set up by the European Commission in 2018. This will not only encourage researchers to actively take this ethical approach to AI but also increase transparency for readers on how ethical the design and implementation process was.

6.3. *Innovating science education with AI through interdisciplinary collaborations*

As stated earlier, the reviewed studies mainly use the new technology to automate existing learning practices. However, to make full use of all the opportunities AI has for school science, we need to start thinking outside the box and start formulating new learning objectives and new teaching materials that were, before the upcoming of AI, unreachable. Instead of thinking about how we can use AI to educate, we have to start asking ourselves how can we educate using AI for a socially just and fair future. Hence, we argue that more interdisciplinary collaborations must be developed between the fields of Computing Science and Educational Research. Computing scientists have the means to develop AI applications, while educational researchers have the theory to make designs pedagogically appropriate, theoretically-sound and socially just.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the European Commission [MAMMOTH].

ORCID

Dagmar Mercedes Heeg http://orcid.org/0000-0002-0929-109X
Lucy Avraamidou http://orcid.org/0000-0001-9693-4438

References

Bosch, N. & D'Mello, S. K. (2021). Automatic Detection of Mind Wandering from Video in the Lab and in the Classroom. *IEEE Transactions on Affective Computing*, *12*(4), 974–988. https://doi.org/10.1109/TAFFC.2019.2908837

Bozkurt, A., Karadeniz, A., Baneres, D., Guerrero-Roldán, A. E., & Rodríguez, M. E. (2021). Artificial intelligence and reflections from educational landscape: A review of ai studies in half a century. *Sustainability*, *13*(2), 800. https://doi.org/10.3390/su13020800

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial intelligence trends in education: A narrative overview. *Procedia Computer Science*, *136*, 16–24. https://doi.org/10.1016/j.procs.2018.08.233

Chu, H. C., Hwang, G. J., & Tsai, C. C. (2010). A knowledge engineering approach to developing mindtools for context-aware ubiquitous learning. *Computers & Education*, *54*(1), 289–297. https://doi.org/10.1016/j.compedu.2009.08.023

Dolenc, K., Abersek, B., & Abersek, M. K. (2015). Online functional literacy, intelligent tutoring systems and science education. *Journal of Baltic Science Education*, *14*(2), 162–171. https://doi.org/10.33225/jbse/15.14.162

European Commission. (2020, 19 February). *White paper on Artificial intelligence – a European approach to excellence and trust*. Author European Commission. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

Feng, S., & Law, N. (2021). Mapping artificial intelligence in education research: A network-based keyword analysis. *International Journal of Artificial Intelligence in Education*, *31*(2), 277–303. https://doi.org/10.1007/s40593-021-00244-4

Gaudioso, E., Montero, M., & Hernandez-Del-Olmo, F. (2012). Supporting teachers in adaptive educational systems through predictive models: A proof of concept. *Expert Systems with Applications*, *39*(1), 621–625. https://doi.org/10.1016/j.eswa.2011.07.052

Gerard, L., Matuk, C., McElhaney, K., & Linn, M. C. (2015). Automated, adaptive guidance for k-12 education. *Educational Research Review*, *15*, 41–58. https://doi.org/10.1016/j.edurev.2015.04.001

Gonzalez, A. J., Hollister, J. R., DeMara, R. F., Leigh, J., Lanman, B., Lee, S.-Y., Parker, S., Walls, C., Parker, J., Wong, J., Barham, C., & Wilder, B. (2017). AI in informal science education: Bringing turing back to life to perform the turing test. *International Journal of Artificial Intelligence in Education: Official Journal of the International AIED Society*, *27*(2), 353–384. https://doi.org/10.1007/s40593-017-0144-1

Guan, C., Mou, J., & Jiang, Z. (2020). Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. *International Journal of Innovation Studies*, *4*(4), 134–147. https://doi.org/10.1016/j.ijis.2020.09.001

Huang, C.-J., Wang, Y.-W., Huang, T.-H., Chen, Y.-C., Chen, H.-M., & Chang, S.-C. (2011). Performance evaluation of an online argumentation learning assistance agent. *Computers & Education*, *57*(1), 1270–1280. https://doi.org/10.1016/j.compedu.2011.01.013

Hutt, S., Krasich, K., Mills, C., Bosch, N., White, S., Brockmole, J. R., & D'Mello, S. K. (2019). Automated gaze-based mind wandering detection during computerized learning in classrooms. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, *29*(4), 821–867. https://doi.org/10.1007/s11257-019-09228-5

Lee, H.-S., Gweon, G.-H., Lord, T., Paessel, N., Pallant, A., & Pryputniewicz, S. (2021). Machine learning-enabled automated feedback: Supporting students' revision of scientific arguments based on data drawn from simulation. *Journal of Science Education and Technology*, *30*(2), 168–192. https://doi.org/10.1007/s10956-020-09889-7

Lee, H.-S., Pallant, A., Pryputniewicz, S., Lord, T., Mulholland, M., & Liu, O. L. (2019). Automated text scoring and real-time adjustable feedback: Supporting revision of scientific arguments involving uncertainty. *Science Education*, *103*(3), 590–622. https://doi.org/10.1002/sce.21504

Maestrales, S., Zhai, X., Touitou, I., Baker, Q., Schneider, B., & Krajcik, J. (2021). Using machine learning to score multi-dimensional assessments of chemistry and physics. *Journal of Science Education and Technology*, *30*(2), 239–254. https://doi.org/10.1007/s10956-020-09895-9

Mao, L., Liu, O. L., Roohr, K., Belur, V., Mulholland, M., Lee, H.-S., & Pallant, A. (2018). Validation of automated scoring for a formative assessment that employs scientific argumentation. *Educational Assessment*, *23*(2), 121–138. https://doi.org/10.1080/10627197.2018.1427570

Mead, C., Buxner, S., Bruce, G., Taylor, W., Semken, S., & Anbar, A. D. (2019). Immersive, interactive virtual field trips promote science learning. *Journal of Geoscience Education*, *67*(2), 131–142. https://doi.org/10.1080/10899995.2019.1565285

Miao, F., Holmes, W., Ronghuai, H., & Zhang H. (2021). AI end education: Guidance for policy-makers (UNESCO), *45*. https://doi.org/10.54675/PCSP7350

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *Annals of Internal Medicine*, *151*(4), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Ntemngwa, C., & Steve, O. J. (2018). The implementation of Integrated Science Technology, engineering and mathematics (STEM) instruction using robotics in the middle school Science classroom. *International Journal of Education in Mathematics, Science and Technology*, 12–40. https://doi.org/10.18404/ijemst.380617

Nye, B. D., Davis, D. M., Rizvi, S. Z., Carr, K., Swartout, W., Thacker, R., & Shaw, K. (2020). Feasibility and usability of MentorPal, a framework for rapid development of virtual mentors. *Journal of Research on Technology in Education*, *53*(1), 21–43. https://doi.org/10.1080/15391523.2020.1771640

Nye, B. D., Davis, D. M., Rizvi, S. Z., Carr, K., Swartout, W., Thacker, R., & Shaw, K. (2021). Feasibility and usability of MentorPal, a framework for rapid development of virtual mentors. *Journal of Research on Technology in Education*, *53*(1), 21–43. https://doi.org/10.1080/15391523.2020.1771640

OECD. (2021). *OECD digital education outlook 2021 : Pushing the frontiers with artificial intelligence, blockchain and robots*. OECD Publishing. https://doi.org/10.1787/589b283f-en

Parsia, J., Leo, B., Al-Emari, U., Sattler, S., & Al-Emari, S. (2020). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, *30*(1), 121–204. https://doi.org/10.1007/s40593-019-00186-y

Ponticorvo, M., Rubinacci, F., Marocco, D., Truglio, F. & Miglino, O. (2020). Educational Robotics to Foster and Assess Social Relations in Students' Groups. *Frontiers in Robotics and AI*, *7*. https://doi.org/10.3389/frobt.2020.00078

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, *12*(1), 22–22. https://doi.org/10.1186/s41039-017-0062-8

San Pedro, M. O., Baker, R. S. & Rodrigo, M. M. (2014). Carelessness and Affect in an Intelligent Tutoring System for Mathematics. *International Journal of Artificial Intelligence in Education*, *24*(2), 189–210. https://doi.org/10.1007/s40593-014-0015-y

Sung, S. H., Li, C., Chen, G., Huang, X., Xie, C., Massicotte, J., & Shen, J. (2021). How does augmented observation facilitate multimodal representational thinking? applying deep learning to decode complex student construct. *Journal of Science Education and Technology*, *30*(2), 210–226. sung. https://doi.org/10.1007/s10956-020-09856-2

Topal, A. D., Eren, C. D., & Geçer, A. K. (2021). Chatbot application in a 5th grade science course. *Education and Information Technologies*, *26*(5), 6241–6265. https://doi.org/10.1007/s10639-021-10627-8

Udandarao, V., Gupta, T., Agarwal, A., & Chakraborty, A. (2021). Inphynet: Leveraging attention-based multitask recurrent networks for multi-label physics text classification. *Knowledge-Based Systems*, *211*, 211. https://doi.org/10.1016/j.knosys.2020.106487

Van Griethuijsen, R. A. L. F., van Eijck, M. W., Haste, H., den Brok, P. J., Skinner, N. C., Mansour, N., Gencer, A. S., & BouJaoude, S. (2015). Global patterns in students' views of science and interest in science. *Research in Science Education*, *45*(4), 581–603. https://doi.org/10.1007/s11165-014-9438-6

Ward, W., Cole, R., Bolaños, D., Buchenroth-Martin, C., Svirsky, E., & Weston, T. (2013). My science tutor: A conversational multimedia virtual tutor. *Journal of Educational Psychology*, *105*(4), 1115–1125. https://doi.org/10.1037/a0031589

Wen, C.-T., Chang, C.-J., Chang, M.-H., Fan Chiang, S.-H., Liu, C.-C., Hwang, F.-K., & Tsai, C.-C. (2018). The learning analytics of model-based learning facilitated by a problem-solving simulation game. *Instructional Science: An International Journal of the Learning Sciences*, *46*(6), 847–867. https://doi.org/10.1007/s11251-018-9461-5

Wu, P.-H., Hwang, G.-J., & Tsai, W.-H. (2013). An expert system-based context-aware ubiquitous learning approach for conducting science learning activities. *Educational Technology & Society*, *16*(4), 217–230.

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, *39*(1), 93–112. https://doi.org/10.1177/0739456X17723971

Yağci, A., & Çevik, M. (2019). Prediction of academic achievements of vocational and technical high school (VTS) students in science courses through artificial neural networks (comparison of Turkey and Malaysia). *Education and Information Technologies: The Official Journal of the IFIP Technical Committee on Education*, *24*(5), 2741–2761. https://doi.org/10.1007/s10639-019-09885-4

Yağci, A., & Çevik, M. (2019). Prediction of academic achievements of vocational and technical high school (VTS) students in science courses through artificial neural networks (comparison of Turkey and Malaysia). *Educ Inf Technol*, *24*(5), 2741–2761. https://doi.org/10.1007/s10639-019-09885-4

Zawacki-Richter, O., Marín, I. V., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education - where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0

Zhai, X., Yin, Y., Pellegrino, J. W., Haudek, K. C., & Shi, L. (2020). Applying machine learning in science assessment: A systematic review. *Studies in Science Education*, *56*(1), 111–151. https://doi.org/10.1080/03057267.2020.1735757

Zhu, M., Lee, H.-S., Wang, T., Liu, O. L., Belur, V., & Pallant, A. (2017). Investigating the impact of automated feedback on students' scientific argumentation. *International Journal of Science Education*, *39*(12), 1648–1668. https://doi.org/10.1080/09500693.2017.1347303

Zulfiani, Z., Suwarna, I. P., & Miranto, S. (2018). Science education adaptive learning system as a computer-based science learning with learning style variations. *Journal of Baltic Science Education*, *17*(4), 711–727. https://doi.org/10.33225/jbse/18.17.711

Zulfiani, Z., Suwarna, I. P., & Miranto, S. (2021). Improving students' academic achievement using the ScEd-ALS Android-based. *International Journal of Instruction*, *14*(2), 735–756. https://doi.org/10.29333/iji.2021.14241a