



# Artificial intelligence in education: Addressing ethical challenges in K-12 settings

Selin Akgun<sup>1</sup> · Christine Greenhow<sup>1</sup>

Received: 9 July 2021 / Accepted: 1 September 2021 / Published online: 22 September 2021  
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

## Abstract

Artificial intelligence (AI) is a field of study that combines the applications of machine learning, algorithm productions, and natural language processing. Applications of AI transform the tools of education. AI has a variety of educational applications, such as personalized learning platforms to promote students' learning, automated assessment systems to aid teachers, and facial recognition systems to generate insights about learners' behaviors. Despite the potential benefits of AI to support students' learning experiences and teachers' practices, the ethical and societal drawbacks of these systems are rarely fully considered in K-12 educational contexts. The ethical challenges of AI in education must be identified and introduced to teachers and students. To address these issues, this paper (1) briefly defines AI through the concepts of machine learning and algorithms; (2) introduces applications of AI in educational settings and benefits of AI systems to support students' learning processes; (3) describes ethical challenges and dilemmas of using AI in education; and (4) addresses the teaching and understanding of AI by providing recommended instructional resources from two providers—i.e., the Massachusetts Institute of Technology's (MIT) Media Lab and Code.org. The article aims to help practitioners reap the benefits and navigate ethical challenges of integrating AI in K-12 classrooms, while also introducing instructional resources that teachers can use to advance K-12 students' understanding of AI and ethics.

**Keywords** Artificial intelligence · K-12 education · Ethics · Teacher education

## 1 Introduction

*“Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last, unless we learn how to avoid the risks.”—Stephen Hawking.*

We may not think about artificial intelligence (AI) on a daily basis, but it is all around us, and we have been using it for years. When we are doing a Google search, reading our emails, getting a doctor's appointment, asking for driving directions, or getting movie and music recommendations, we are constantly using the applications of AI and its assistance in our lives. This need for assistance and our dependence on AI systems has become even more apparent during the COVID-19 pandemic. The growing impact and dominance

of AI systems reveals itself in healthcare, education, communications, transportation, agriculture, and more. It is almost impossible to live in a modern society without encountering applications powered by AI [10, 32].

Artificial intelligence (AI) can be defined briefly as the branch of computer science that deals with the simulation of intelligent behavior in computers and their capacity to mimic, and ideally improve, human behavior [43]. AI dominates the fields of science, engineering, and technology, but also is present in education through machine-learning systems and algorithm productions [43]. For instance, AI has a variety of algorithmic applications in education, such as personalized learning systems to promote students' learning, automated assessment systems to support teachers in evaluating what students know, and facial recognition systems to provide insights about learners' behaviors [49]. Besides these platforms, algorithm systems are prominent in education through different social media outlets, such as social network sites, microblogging systems, and mobile applications. Social media are increasingly integrated into K-12 education [7] and subordinate learners'

✉ Selin Akgun  
akgunsel@msu.edu

Christine Greenhow  
greenhow@msu.edu

<sup>1</sup> Michigan State University, East Lansing, MI, USA

activities to intelligent algorithm systems [17]. Here, we use the American term “K–12 education” to refer to students’ education in kindergarten (K) (ages 5–6) through 12th grade (ages 17–18) in the United States, which is similar to primary and secondary education or pre-college level schooling in other countries. These AI systems can increase the capacity of K-12 educational systems and support the social and cognitive development of students and teachers [55, 8]. More specifically, applications of AI can support instruction in mixed-ability classrooms; while personalized learning systems provide students with detailed and timely feedback about their writing products, automated assessment systems support teachers by freeing them from excessive workloads [26, 42].

Despite the benefits of AI applications for education, they pose societal and ethical drawbacks. As the famous scientist, Stephen Hawking, pointed out that weighing these risks is vital for the future of humanity. Therefore, it is critical to take action toward addressing them. The biggest risks of integrating these algorithms in K-12 contexts are: (a) perpetuating existing systemic bias and discrimination, (b) perpetuating unfairness for students from mostly disadvantaged and marginalized groups, and (c) amplifying racism, sexism, xenophobia, and other forms of injustice and inequity [40]. These algorithms do not occur in a vacuum; rather, they shape and are shaped by ever-evolving cultural, social, institutional and political forces and structures [33, 34]. As academics, scientists, and citizens, we have a responsibility to educate teachers and students to recognize the ethical challenges and implications of algorithm use. To create a future generation where an inclusive and diverse citizenry can participate in the development of the future of AI, we need to develop opportunities for K-12 students and teachers to learn about AI via AI- and ethics-based curricula and professional development [2, 58]

Toward this end, the existing literature provides little guidance and contains a limited number of studies that focus on supporting K-12 students and teachers’ understanding of social, cultural, and ethical implications of AI [2]. Most studies reflect university students’ engagement with ethical ideas about algorithmic bias, but few addresses how to promote students’ understanding of AI and ethics in K-12 settings. Therefore, this article: (a) synthesizes ethical issues surrounding AI in education as identified in the educational literature, (b) reflects on different approaches and curriculum materials available for teaching students about AI and ethics (i.e., featuring materials from the MIT Media Lab and Code.org), and (c) articulates future directions for research and recommendations for practitioners seeking to navigate AI and ethics in K-12 settings.

Next, we briefly define the notion of artificial intelligence (AI) and its applications through machine-learning and algorithm systems. As educational and educational technology scholars working in the United States, and at the risk of oversimplifying, we provide only a brief definition of AI below, and recognize that definitions of AI are complex, multidimensional, and contested in the literature [9, 16, 38]; an in-depth discussion of these complexities, however, is beyond the scope of this paper. Second, we describe in more detail five applications of AI in education, outlining their potential benefits for educators and students. Third, we describe the ethical challenges they raise by posing the question: “how and in what ways do algorithms manipulate us?” Fourth, we explain how to support students’ learning about AI and ethics through different curriculum materials and teaching practices in K-12 settings. Our goal here is to provide strategies for practitioners to reap the benefits while navigating the ethical challenges. We acknowledge that in centering this work within U.S. education, we highlight certain ethical issues that educators in other parts of the world may see as less prominent. For example, the European Union (EU) has highlighted ethical concerns and implications of AI, emphasized privacy protection, surveillance, and non-discrimination as primary areas of interest, and provided guidelines on how trustworthy AI should be [3, 15, 23]. Finally, we reflect on future directions for educational and other research that could support K-12 teachers and students in reaping the benefits while mitigating the drawbacks of AI in education.

## 2 Definition and applications of artificial intelligence

The pursuit of creating intelligent machines that replicate human behavior has accelerated with the realization of artificial intelligence. With the latest advancements in computer science, a proliferation of definitions and explanations of what counts as AI systems has emerged. For instance, AI has been defined as “the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings” [49]. This particular definition highlights the mimicry of human behavior and consciousness. Furthermore, AI has been defined as “the combination of cognitive automation, machine learning, reasoning, hypothesis generation and analysis, natural language processing, and intentional algorithm mutation producing insights and analytics at or above human capability” [31]. This definition incorporates the different sub-fields of AI together and

underlines their function while reaching at or above human capability.

Combining these definitions, artificial intelligence can be described as the *technology that builds systems to think and act like humans with the ability of achieving goals*. AI is mainly known through different applications and advanced computer programs, such as recommender systems (e.g., YouTube, Netflix), personal assistants (e.g., Apple’s Siri), facial recognition systems (e.g., Facebook’s face detection in photographs), and learning apps (e.g., Duolingo) [32]. To build on these programs, different sub-fields of AI have been used in a diverse range of applications. Evolutionary algorithms and machine learning are most relevant to AI in K-12 education.

## 2.1 Algorithms

Algorithms are the core elements of AI. The history of AI is closely connected to the development of sophisticated and evolutionary algorithms. An algorithm is a set of rules or instructions that is to be followed by computers in problem-solving operations to achieve an intended end goal. In essence, all computer programs are algorithms. They involve thousands of lines of codes which represent mathematical instructions that the computer follows to solve the intended problems (e.g., as computing numerical calculation, processing an image, and grammar-checking in an essay). AI algorithms are applied to fields that we might think of as essentially human behavior—such as speech and face recognition, visual perception, learning, and decision-making and learning. In that way, algorithms can provide instructions for almost any AI system and application we can conceive [27].

## 2.2 Machine learning

Machine learning is derived from statistical learning methods and uses data and algorithms to perform tasks which are typically performed by humans [43]. Machine learning is about making computers act or perform without being given any line-by-line step [29]. The working mechanism of machine learning is the learning model’s exposure to ample amounts of quality data [41]. Machine-learning algorithms first analyze the data to determine patterns and to build a model and then predict future values through these models. In other words, machine learning can be considered a three-step process. First, it analyzes and gathers the data, and then, it builds a model to excel for different tasks, and finally, it undertakes the action and produces the desired results successfully without human intervention [29, 56]. The widely known AI applications such as recommender or facial recognition systems have all been made possible through the working principles of machine learning.

## 3 Benefits of AI applications in education

Personalized learning systems, automated assessments, facial recognition systems, chatbots (social media sites), and predictive analytics tools are being deployed increasingly in K-12 educational settings; they are powered by machine-learning systems and algorithms [29]. These applications of AI have shown promise to support teachers and students in various ways: (a) providing instruction in mixed-ability classrooms, (b) providing students with detailed and timely feedback on their writing products, (c) freeing teachers from the burden of possessing all knowledge and giving them more room to support their students while they are observing, discussing, and gathering information in their collaborative knowledge-building processes [26, 50]. Below, we outline benefits of each of these educational applications in the K-12 setting before turning to a synthesis of their ethical challenges and drawbacks.

### 3.1 Personalized learning systems

Personalized learning systems, also known as adaptive learning platforms or intelligent tutoring systems, are one of the most common and valuable applications of AI to support students and teachers. They provide students access to different learning materials based on their individual learning needs and subjects [55]. For example, rather than practicing chemistry on a worksheet or reading a textbook, students may use an adaptive and interactive multimedia version of the course content [39]. Comparing students’ scores on researcher-developed or standardized tests, research shows that the instruction based on personalized learning systems resulted in higher test scores than traditional teacher-led instruction [36]. Microsoft’s recent report (2018) of over 2000 students and teachers from Singapore, the U.S., the UK, and Canada shows that AI supports students’ learning progressions. These platforms promise to identify gaps in students’ prior knowledge by accommodating learning tools and materials to support students’ growth. These systems generate models of learners using their knowledge and cognition; however, the existing platforms do not yet provide models for learners’ social, emotional, and motivational states [28]. Considering the shift to remote K-12 education during the COVID-19 pandemic, personalized learning systems offer a promising form of distance learning that could reshape K-12 instruction for the future [35].

### 3.2 Automated assessment systems

Automated assessment systems are becoming one of the most prominent and promising applications of machine

learning in K-12 education [42]. These scoring algorithm systems are being developed to meet the need for scoring students' writing, exams and assignments, and tasks usually performed by the teacher. Assessment algorithms can provide course support and management tools to lessen teachers' workload, as well as extend their capacity and productivity. Ideally, these systems can provide levels of support to students, as their essays can be graded quickly [55]. Providers of the biggest open online courses such as Coursera and EdX have integrated automated scoring engines into their learning platforms to assess the writings of hundreds of students [42]. On the other hand, a tool called "Gradescope" has been used by over 500 universities to develop and streamline scoring and assessment [12]. By flagging the wrong answers and marking the correct ones, the tool supports instructors by eliminating their manual grading time and effort. Thus, automated assessment systems deal very differently with marking and giving feedback to essays compared to numeric assessments which analyze right or wrong answers on the test. Overall, these scoring systems have the potential to deal with the complexities of the teaching context and support students' learning process by providing them with feedback and guidance to improve and revise their writing.

### 3.3 Facial recognition systems and predictive analytics

Facial recognition software is used to capture and monitor students' facial expressions. These systems provide insights about students' behaviors during learning processes and allow teachers to take action or intervene, which, in turn, helps teachers develop learner-centered practices and increase student's engagement [55]. Predictive analytics algorithm systems are mainly used to identify and detect patterns about learners based on statistical analysis. For example, these analytics can be used to detect university students who are at risk of failing or not completing a course. Through these identifications, instructors can intervene and get students the help they need [55].

### 3.4 Social networking sites and chatbots

Social networking sites (SNSs) connect students and teachers through social media outlets. Researchers have emphasized the importance of using SNSs (such as Facebook) to expand learning opportunities beyond the classroom, monitor students' well-being, and deepen student–teacher relations [5]. Different scholars have examined the role of social media in education, describing its impact on student and teacher learning and scholarly communication [6]. They

point out that the integration of social media can foster students' active learning, collaboration skills, and connections with communities beyond the classroom [6]. Chatbots also take place in social media outlets through different AI systems [21]. They are also known as dialogue systems or conversational agents [26, 52]. Chatbots are helpful in terms of their ability to respond naturally with a conversational tone. For instance, a text-based chatbot system called "Pounce" was used at Georgia State University to help students through the registration and admission process, as well as financial aid and other administrative tasks [7].

In summary, applications of AI can positively impact students' and teachers' educational experiences and help them address instructional challenges and concerns. On the other hand, AI cannot be a substitute for human interaction [22, 47]. Students have a wide range of learning styles and needs. Although AI can be a time-saving and cognitive aide for teachers, it is but one tool in the teachers' toolkit. Therefore, it is critical for teachers and students to understand the limits, potential risks, and ethical drawbacks of AI applications in education if they are to reap the benefits of AI and minimize the costs [11].

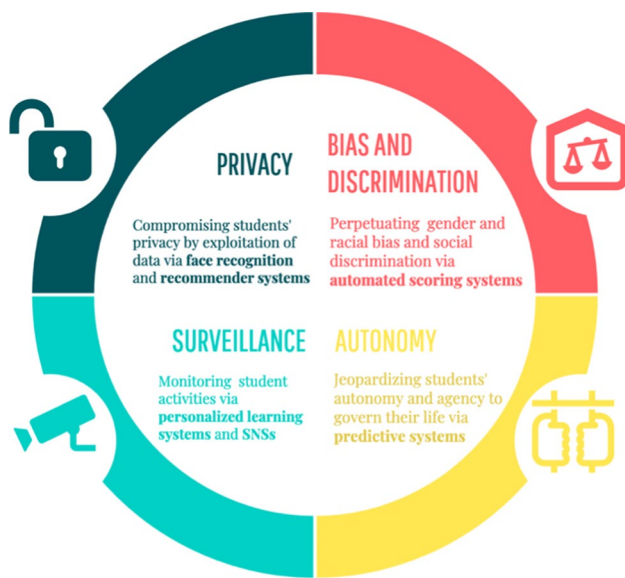
## 4 Ethical concerns and potential risks of AI applications in education

The ethical challenges and risks posed by AI systems seemingly run counter to marketing efforts that present algorithms to the public as if they are objective and value-neutral tools. In essence, algorithms reflect the values of their builders who hold positions of power [26]. Whenever people create algorithms, they also create a set of data that represent society's historical and systemic biases, which ultimately transform into algorithmic bias. Even though the bias is embedded into the algorithmic model with no explicit intention, we can see various gender and racial biases in different AI-based platforms [54].

Considering the different forms of bias and ethical challenges of AI applications in K-12 settings, we will focus on problems of privacy, surveillance, autonomy, bias, and discrimination (see Fig. 1). However, it is important to acknowledge that educators will have different ethical concerns and challenges depending on their students' grade and age of development. Where strategies and resources are recommended, we indicate the age and/or grade level of student(s) they are targeting (Fig. 2).

One of the biggest ethical issues surrounding the use of AI in K-12 education relates to the *privacy* concerns





**Fig. 1** Potential ethical and societal risks of AI applications in education



**Fig. 2** Student work from the activity of “Youtube Redesign” (MIT Media Lab, AI and Ethics Curriculum, p.1, [45])

of students and teachers [47, 49, 54]. Privacy violations mainly occur as people expose an excessive amount of personal information in online platforms. Although existing legislation and standards exist to protect sensitive personal data, AI-based tech companies' violations with respect to data access and security increase people's privacy concerns [42, 54]. To address these concerns, AI systems ask for users' consent to access their personal data. Although consent requests are designed to be protective measures and to help alleviate privacy concerns, many individuals give their consent without knowing or considering the extent of the information (metadata) they are sharing, such as the language spoken, racial identity, biographical data, and location [49]. Such uninformed sharing in effect

undermines human agency and privacy. In other words, people's agency diminishes as AI systems reduce introspective and independent thought [55]. Relatedly, scholars have raised the ethical issue of forcing students and parents to use these algorithms as part of their education even if they explicitly agree to give up privacy [14, 48]. They really have no choice if these systems are required by public schools.

Another ethical concern surrounding the use of AI in K-12 education is *surveillance* or tracking systems which gather detailed information about the actions and preferences of students and teachers. Through algorithms and machine-learning models, AI tracking systems not only necessitate monitoring of activities but also determine the future preferences and actions of their users [47]. Surveillance mechanisms can be embedded into AI's predictive systems to foresee students' learning performances, strengths, weaknesses, and learning patterns. For instance, research suggests that teachers who use social networking sites (SNSs) for pedagogical purposes encounter a number of problems, such as concerns in relation to boundaries of privacy, friendship authority, as well as responsibility and availability [5]. While monitoring and patrolling students' actions might be considered part of a teacher's responsibility and a pedagogical tool to intervene in dangerous online cases (such as cyber-bullying or exposure to sexual content), such actions can also be seen as surveillance systems which are problematic in terms of threatening students' privacy. Monitoring and tracking students' online conversations and actions also may limit their participation in the learning event and make them feel unsafe to take ownership for their ideas. How can students feel secure and safe, if they know that AI systems are used for surveilling and policing their thoughts and actions? [49].

Problems also emerge when surveillance systems trigger issues related to *autonomy*, more specifically, the person's ability to act on her or his own interest and values. Predictive systems which are powered by algorithms jeopardize students and teachers' autonomy and their ability to govern their own life [46, 47]. Use of algorithms to make predictions about individuals' actions based on their information raise questions about fairness and self-freedom [19]. Therefore, the risks of predictive analysis also include the perpetuation of existing bias and prejudices of social discrimination and stratification [42].

Finally, *bias and discrimination* are critical concerns in debates of AI ethics in K-12 education [6]. In AI platforms, the existing power structures and biases are embedded into machine-learning models [6]. Gender bias is one of the most apparent forms of this problem, as the bias is revealed when students in language learning courses use AI to translate between a gender-specific language and one that is less-so. For example, while Google Translate

translated the Turkish equivalent of “*She/he is a nurse*” into the feminine form, it also translated the Turkish equivalent of “*She/he is a doctor*” into the masculine form [33]. This shows how AI models in language translation carry the societal biases and gender-specific stereotypes in the data [40]. Similarly, a number of problematic cases of racial bias are also associated with AI’s facial recognition systems. Research shows that facial recognition software has improperly misidentified a number of African American and Latino American people as convicted felons [42].

Additionally, biased decision-making algorithms reveal themselves throughout AI applications in K-12 education: personalized learning, automated assessment, SNSs, and predictive systems in education. Although the main promise of machine-learning models is increased accuracy and objectivity, current incidents have revealed the contrary. For instance, England’s A-level and GCSE secondary level examinations were cancelled due to the pandemic in the summer of 2020 [1, 57]. An alternative assessment method was implemented to determine the qualification grades of students. The grade standardization algorithm was produced by the regulator Ofqual. With the assessment of Ofqual’s algorithm based on schools’ previous examination results, thousands of students were shocked to receive unexpectedly low grades. Although a full discussion of the incident is beyond the scope of this article [51] it revealed how the score distribution favored students who attended private or independent schools, while students from underrepresented groups were hit hardest. Unfortunately, automated assessment algorithms have the potential to reconstruct unfair and inconsistent results by disrupting student’s final scores and future careers [53].

## 5 Teaching and understanding AI and ethics in educational settings

These ethical concerns suggest an urgent need to introduce students and teachers to the ethical challenges surrounding AI applications in K-12 education and how to navigate them. To meet this need, different research groups and nonprofit organizations offer a number of open-access resources based on AI and ethics. They provide instructional materials for students and teachers, such as lesson plans and hands-on activities, and professional learning materials for educators, such as open virtual learning sessions. Below, we describe and evaluate three resources: “AI and Ethics” curriculum and “AI and Data Privacy” workshop from the Massachusetts Institute of Technology (MIT) Media Lab as well as Code.org’s “AI and Oceans” activity. For readers who seek to investigate additional approaches and resources for K-12 level AI and ethics

interaction, see: (a) The Chinese University of Hong Kong (CUHK)’s AI for the Future Project (AI4Future) [18]; (b) IBM’s Educator’s AI Classroom Kit [30], Google’s Teachable Machine [25], UK-based nonprofit organization Apps for Good [4], and Machine Learning for Kids [37].

### 5.1 “AI and Ethics Curriculum” for middle school students by MIT Media Lab

The MIT Media Lab team offers an open-access curriculum on AI and ethics for middle school students and teachers. Through a series of lesson plans and hand-on activities, teachers are guided to support students’ learning of the technical terminology of AI systems as well as the ethical and societal implications of AI [2]. The curriculum includes various lessons tied to learning objectives. One of the main learning goals is to introduce students to basic components of AI through algorithms, datasets, and supervised machine-learning systems all while underlining the problem of algorithmic bias [45]. For instance, in the activity “*AI Bingo*”, students are given bingo cards with various AI systems, such as online search engine, customer service bot, and weather app. Students work with their partners collaboratively on these AI systems. In their AI Bingo chart, students try to identify what prediction the selected AI system makes and what dataset it uses. In that way, they become more familiar with the notions of dataset and prediction in the context of AI systems [45].

In the second investigation, “*Algorithms as Opinions*”, students think about algorithms as recipes, which are created by set of instructions that modify an input to produce an output [45]. Initially, students are asked to write an algorithm to make the “best” jelly sandwich and peanut butter. They explore what it means to be “best” and see how their opinions of best in their recipes are reflected in their algorithms. In this way, students are able to figure out that algorithms can have various motives and goals. Following this activity, students work on the “*Ethical Matrix*”, building on the idea of the algorithms as opinions [45]. During this investigation, students first refer back to their developed algorithms through their “best” jelly sandwich and peanut butter. They discuss what counts as the “best” sandwich for themselves (most healthy, practical, delicious, etc.). Then, through their ethical matrix (chart), students identify different stakeholders (such as their parents, teacher, or doctor) who care about their peanut butter and jelly sandwich algorithm. In this way, the values and opinions of those stakeholders also are embedded in the algorithm. Students fill out an ethical matrix and look for where those values conflict or overlap with each other. This matrix is a great tool for students to recognize different stakeholders in a system or society and how they are able to build and utilize the values of the algorithms in an ethical matrix.

The final investigation which teaches about the biased nature of algorithms is *“Learning and Algorithmic Bias”* [45]. During the investigation, students think further about the concept of classification. Using Google’s Teachable Machine tool [2], students explore the supervised machine-learning systems. Students train a cat–dog classifier using two different datasets. While the first dataset reflects the cats as the over-represented group, the second dataset indicates the equal and diverse representation between dogs and cats [2]. Using these datasets, students compare the accuracy between the classifiers and then discuss which dataset and outcome are fairer. This activity leads students into a discussion about the occurrence of bias in facial recognition algorithms and systems [2].

In the rest of the curriculum, similar to the AI Bingo investigation, students work with their partners to determine the various forms of AI systems in the YouTube platform (such as its recommender algorithm and advertisement matching algorithm). Through the investigation of *“YouTube Redesign”*, students redesign YouTube’s recommender system. They first identify stakeholders and their values in the system, and then use an ethical matrix to reflect on the goals of their YouTube’s recommendation algorithm [45]. Finally, through the activity of *“YouTube Socratic Seminar”*, students read an abridged version of *Wall Street Journal* article by participating in a Socratic seminar. The article was edited to shorten the text and to provide more accessible language for middle school students. They discuss which stakeholders were most influential or significant in proposing changes in the YouTube Kids app and whether or not technologies like auto play should ever exist. During their discussion, students engage with the questions of: “Which stakeholder is making the most change or has the most power?”, “Have you ever seen an inappropriate piece of content on YouTube? What did you do?” [45].

Overall, the MIT Media Lab’s AI and Ethics curriculum is a high quality, open-access resource with which teachers can introduce middle school students to the risks and ethical implications of AI systems. The investigations described above involve students in collaborative, critical thinking activities that force them to wrestle with issues of bias and discrimination in AI, as well as surveillance and autonomy through the predictive systems and algorithmic bias.

## 5.2 “AI and Data Privacy” workshop series for K-9 students by MIT Media Lab

Another quality resource from the MIT Media Lab’s Personal Robots Group is a workshop series designed to teach students (between the ages 7 and 14) about data privacy and introduce them to designing and prototyping data privacy features. The group has made the content, materials, worksheets, and activities of the workshop series into an

open-access online document, freely available to teachers [44].

The first workshop in the series is *“Mystery YouTube Viewer: A lesson on Data Privacy”*. During the workshop, students engage with the question of what privacy and data mean [44]. They observe YouTube’s home page from the perspective of a mystery user. Using the clues from the videos, students make predictions about what the characters in the videos might look like or where they might live. In a way, students imitate YouTube algorithms’ prediction mode about the characters. Engaging with these questions and observations, students think further about why privacy and boundaries are important and how each algorithm will interpret us differently based on who creates the algorithm itself.

The second workshop in the series is *“Designing ads with transparency: A creative workshop”*. Through this workshop, students are able to think further about the meaning, aim, and impact of advertising and the role of advertisements in our lives [44]. Students collaboratively create an advertisement using an everyday object. The objective is to make the advertisement as “transparent” as possible. To do that, students learn about notions of malware and adware, as well as the components of YouTube advertisements (such as sponsored labels, logos, news sections, etc.). By the end of the workshop, students design their ads as a poster, and they share with their peers.

The final workshop in MIT’s AI and data privacy series is *“Designing Privacy in Social Media Platforms”*. This workshop is designed to teach students about YouTube, design, civics, and data privacy [44]. During the workshop, students create their own designs to solve one of the biggest challenges of the digital era: problems associated with online consent. The workshop allows students to learn more about the privacy laws and how they impact youth in terms of media consumption. Students consider YouTube within the lenses of the Children’s Online Privacy Protections Rule (COPPA). In this way, students reflect on one of the components of the legislation: how might students get parental permission (or verifiable consent)?

Such workshop resources seem promising in helping educate students and teachers about the ethical challenges of AI in education. Specifically, social media such as YouTube are widely used as a teaching and learning tool within K-12 classrooms and beyond them, in students’ everyday lives. These workshop resources may facilitate teachers’ and students’ knowledge of data privacy issues and support them in thinking further about how to protect privacy online. Moreover, educators seeking to implement such resources should consider engaging students in the larger question: who should own one’s data? Teaching students the underlying reasons for laws and facilitating debate on the extent to which they are just or not could help get at this question.

### 5.3 Investigation of “AI for Oceans” by Code.org

A third recommended resource for K-12 educators trying to navigate the ethical challenges of AI with their students comes from Code.org, a nonprofit organization focused on expanding students’ participation in computer science. Sponsored by Microsoft, Facebook, Amazon, Google, and other tech companies, Code.org aims to provide opportunities for K-12 students to learn about AI and machine-learning systems [20]. To support students (grades 3–12) in learning about AI, algorithms, machine learning, and bias, the organization offers an activity called “*AI for Oceans*”, where students are able to train their machine-learning models.

The activity is provided as an open-access tutorial for teachers to help their students explore how to train, model and classify data, as well as to understand how human bias plays a role in machine-learning systems. During the activity, students first classify the objects as either “fish” or “not fish” in an attempt to remove trash from the ocean. Then, they expand their training data set by including other sea creatures that belong underwater. Throughout the activity, students are also able to watch and interact with a number of visuals and video tutorials. With the support of their teachers, they discuss machine learning, steps and influences of training data, as well as the formation and risks of biased data [20].

## 6 Future directions for research and teaching on AI and ethics

In this paper, we provided an overview of the possibilities and potential ethical and societal risks of AI integration in education. To help address these risks, we highlighted several instructional strategies and resources for practitioners seeking to integrate AI applications in K-12 education and/or instruct students about the ethical issues they pose. These instructional materials have the potential to help students and teachers reap the powerful benefits of AI while navigating ethical challenges especially related to privacy concerns and bias. Existing research on AI in education provides insight on supporting students’ understanding and use of AI [2, 13]; however, research on how to develop K-12 teachers’ instructional practices regarding AI and ethics is still in its infancy.

Moreover, current resources, as demonstrated above, mainly address privacy and bias-related ethical and societal concerns of AI. Conducting more exploratory and critical research on teachers’ and students’ surveillance and autonomy concerns will be important to designing future resources. In addition, curriculum developers and workshop designers might consider centering culturally relevant and

responsive pedagogies (by focusing on students’ funds of knowledge, family background, and cultural experiences) while creating instructional materials that address surveillance, privacy, autonomy, and bias. In such student-centered learning environments, students voice their own cultural and contextual experiences while trying to critique and disrupt existing power structures and cultivate their social awareness [24, 36].

Finally, as scholars in teacher education and educational technology, we believe that educating future generations of diverse citizens to participate in the ethical use and development of AI will require more professional development for K-12 teachers (both pre-service and in-service). For instance, through sustained professional learning sessions, teachers could engage with suggested curriculum resources and teaching strategies as well as build a community of practice where they can share and critically reflect on their experiences with other teachers. Further research on such reflective teaching practices and students’ sense-making processes in relation to AI and ethics lessons will be essential to developing curriculum materials and pedagogies relevant to a broad base of educators and students.

**Funding** This work was supported by the Graduate School at Michigan State University, College of Education Summer Research Fellowship.

**Data availability** Not applicable.

**Code availability** Not applicable.

### Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

### References

1. Adams, R., McIntyre, N.: England A-level downgrades hit pupils from disadvantaged areas hardest. <https://www.theguardian.com/education/2020/aug/13/england-a-level-downgrades-hit-pupils-from-disadvantaged-areas-hardest> (2020). Accessed 10 September 2020
2. Ali, S. A., Payne, B. H., Williams, R., Park, H. W., Breazeal, C.: Constructionism, ethics, and creativity: developing primary and middle school artificial intelligence education. Paper presented at International Workshop on Education in Artificial Intelligence (EDUAI). Palo Alto, CA, USA. (2019)
3. Almeida, D., Shmarko, K., Lomas, E.: The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of US, EU, and UK regulatory frameworks. *AI Ethics* (2021). <https://doi.org/10.1007/s43681-021-00077-w>
4. Apps for Good: <https://www.appsforgood.org/about> Accessed 28 August 2021



5. Asterhan, C.S.C., Rosenberg, H.: The promise, reality and dilemmas of secondary school teacher–student interactions in Facebook: the teacher perspective. *Comput. Educ.* **85**, 134–148 (2015)
6. Krutka, D., Manca, S., Galvin, S., Greenhow, C., Koehler, M., Askari, E.: Teaching “against” social media: confronting problems of profit in the curriculum. *Teachers College Record* **121**(14), 1–42 (2019)
7. Greenhow, C., Galvin, S., Brandon, D., Askari, E.: A decade of research on K-12 teaching and teacher learning with social media: insights on the state-of-the-field. *Teachers College Record* **122**(6), 1–7 (2020)
8. Greenhow, C., Galvin, S., Staudt Willet, K.B.: Inquiring tweets want to know: #Edchat supports for #remoteteaching during COVID-19. *British Journal of Educational Technology.* **1–21**, (2021)
9. Barr, A., Feigenbaum, E.A.: *Handbook of artificial intelligence*, vol. 1 and 2. Kaufmann, Los Altos (1981)
10. Bencid, R.: Rise of the machines. *Univ. Bus.* **21**(10), 53–54 (2018)
11. Blank, G., Bolsover, G., Dubois, E.: *A new privacy paradox: young people and privacy on social network site*. Oxford Internet Institute (2014)
12. Blumenstyk, G.: Can artificial intelligence make teaching more personal? *The Chronicle of Higher Education*. <https://www.chronicle.com/article/Can-Artificial-Intelligence/243023> (2018)
13. Borenstein, J., Howard, A.: Emerging challenges in AI and the need for AI ethics education. *AI Ethics* **1**, 61–65 (2020)
14. Bulger, M.: *Personalized learning: the conversations we’re not having*. Data and Society Research Institute. <https://datasociety.net/library/personalized-learning-the-conversations-were-not-having/> (2016)
15. Cath, C., Wachter, S., Mittelstadt, B., Taddei, M., Floridi, L.: Artificial Intelligence and the ‘good society’: the US, EU, and UK approach. *Sci. Eng. Ethics* **24**, 505–528 (2018)
16. Chaudhry, M.A., Kazim, E.: *Artificial intelligence in education (AIED): a high-level academic and industry note 2021*. *AI Ethics* (2021). <https://doi.org/10.1007/s43681-021-00074-z>
17. Cheney-Lippold, J.: *We are data: algorithms and the making of our digital selves*. New York University Press, New York (2017)
18. Chiu, T.K.F., Meng, H., Chai, C.S., King, I., Wong, S., Yam, Y.: Creation and evaluation of a pre-tertiary artificial intelligence (AI) curriculum. *IEEE Trans. Educ.* (2021). <https://doi.org/10.1109/TE.2021.3085878>
19. Citron, D.K., Pasquale, F.A.: The scored society: due process for automated predictions. *Wash. Law Rev.* **89**, 1–33 (2014)
20. Code.org.: *AI for Oceans*. <https://code.org/oceans> (2020)
21. Dignum, V.: Ethics in artificial intelligence: introduction to the special issue. *Ethics Inf. Technol.* **20**, 1–3 (2018)
22. Dishon, G.: New data, old tensions: big data, personalized learning, and the challenges of progressive education. *Theory Res. Educ.* **15**(3), 272–289 (2017)
23. Tiple, Vasile, Recommendations on the European Commission’s WHITE PAPER on Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 final (the ‘AI White Paper’) (2020). <https://doi.org/10.2139/ssrn.3706099>
24. Gay, G.: *Culturally responsive teaching: theory, research, and practice*. Teachers College Press, New York (2010)
25. Google Teachable Machine: <https://teachablemachine.withgoogle.com/> Accessed 28 August 2021
26. Hrastinski, S., Olofsson, A.D., Arkenback, C., Ekström, S., Ericsson, E., Fransson, G., Jaldemark, J., Ryberg, T., Öberg, L., Fuentes, A., Gustafsson, U., Humble, N., Mozelius, P., Sundgren, M., Utterberg, M.: Critical imaginaries and reflections on artificial intelligence and robots in postdigital K-12 education. *Postdigit. Sci. Educ.* **1**, 427–445 (2019)
27. Henderson, P., Sinha, K., Angelard-Gontier, N., Ke, N. R., Fried, G., Lowe, R., Pineau, J.: Ethical challenges in data-driven dialogue systems. In: *Proceedings of AAAI/ACM Conference on AI Ethics and Society (AIES-18)*, New Orleans, Louisiana, USA. (2000)
28. Herder, E., Sosnovsky, S., Dimitrova, V.: Adaptive intelligent learning environments. In: Duval, Erik, Sharples, Mike, Sutherland, Rosamund (eds.) *Technology enhanced learning*, pp. 109–114. Springer International Publishing, Cham (2017)
29. Holmes, W., Bialik, M., Fadel, C.: *Artificial intelligence in education: promises and implications for teaching and learning*. Center for Curriculum Redesign, Boston (2019)
30. IBM the Educator’s AI in the Classroom Toolkit: <https://docs.google.com/document/d/1Zqi74ejYYLAAEjFBjIYuXUCOMQI56R6K0ZsvQbDmOA/edit> Accessed 28 August 2021
31. IEEE Corporate Advisory Group (CAG): *IEEE guide for terms and concepts in intelligent process automation*. The Institute of Electrical and Electronics Engineers Standards Association. 1–16 (2017). <https://ieeexplore.ieee.org/iel7/8070669/8070670/08070671.pdf>
32. Iman, M., Arabnia, H. R., Branchinst, R. M.: Pathways to artificial general intelligence: a brief overview of developments and ethical issues via artificial intelligence, machine learning, deep learning, and data science. Paper presented at International Conference on Artificial Intelligence (ICAI). Las Vegas, Nevada, USA. (2020)
33. Johnson, M.: A scalable approach to reducing gender bias in Google translate. <https://ai.googleblog.com/2020/04/a-scalable-approach-to-reducing-gender.html> Accessed 26 March 2021
34. Ko, A.J., Oleson, A., Ryan, N., Register, Y., Xie, B., Tari, M., Davidson, M., Druga, S., Loksa, D.: It is time for more critical CS education. *Commun. ACM* **63**(11), 31–33 (2020)
35. Krueger, N.: Artificial intelligence has infiltrated our lives. Can it improve learning? *International Society for Technology in Education (ISTE)*. (2017).
36. Ladson-Billings, G.: Toward a theory of culturally relevant pedagogy. *Am. Educ. Res. J.* **32**(3), 465–491 (1995)
37. Machine Learning for Kids: <https://machinelearningforkids.co.uk#!/links>. Accessed 28 August 2021
38. McCarthy, J.: What is artificial intelligence? <http://jmc.stanford.edu/artificial-intelligence/what-is-ai/> Accessed 28 August 2021
39. McMurtrie, B.: How artificial intelligence is changing teaching. *The Chronicle of Higher Education*. <https://www.chronicle.com/article/How-Artificial-Intelligence-Is/244231> (2018)
40. Miller, F.A., Katz, J.H., Gans, R.: The OD imperative to add inclusion to the algorithms of artificial intelligence. *OD Practitioner.* **5**(1), 6–12 (2018)
41. Mohri, M., Rostamizadeh, A., Talwalkar, A.: *Foundations of machine learning*. MIT Press, Cambridge (2012)
42. Murphy, R. F.: Artificial intelligence applications to support k–12 teachers and teaching: a review of promising applications, challenges, and risks. *Perspective.* 1–20 (2019). <https://doi.org/10.7249/PE315>
43. Naqvi, A.: *Artificial intelligence for audit, forensic accounting, and valuation: a strategic perspective*. Wiley (2020)
44. Nguyen, S., DiPaola, D.: AI & data privacy activities for k-9 students. MIT Media Lab. <https://www.media.mit.edu/projects/data-privacy-design-for-youth/overview/> (2020)
45. Payne, B. H.: An ethics of artificial intelligence curriculum for middle school students. MIT Media Lab. <https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/> (2019)
46. Piano, S.L.: Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanit. Soc. Sci. Commun.* **7**(9), 1–7 (2020)

47. Regan, P.M., Jesse, J.: Ethical challenges of edtech, big data and personalized learning: twenty-first century student sorting and tracking. *Ethics Inf. Technol.* **21**, 167–179 (2019)
48. Regan, P. M., Steeves, V.: Education, privacy, and big data algorithms: taking the persons out of personalized learning. *First Monday*, **24**(11), (2019)
49. Remian, D.: Augmenting education: ethical considerations for incorporating artificial intelligence in education (Unpublished master's thesis). University of Massachusetts, Boston (2019)
50. Roll, I., Wylie, R.: Evolution and revolution in artificial intelligence in education. *Int. J. Artif. Intell. Educ.* **26**, 582–599 (2016)
51. Smith, H.: Algorithmic bias: should students pay the price? *AI Soc.* **35**, 1077–1078 (2020)
52. Smutny, P., Schreiberova, P.: Chatbots for learning: a review of educational chatbots for the Facebook Messenger. *Comput. Educ.* **151**, 1–11 (2020)
53. Specia, M.: Parents, Students and Teachers Give Britain a Failing Grade Over Exam Results. *The New York Times*, (2020). <https://www.nytimes.com/2020/08/14/world/europe/england-a-level-results.html>
54. Stahl, B.C., Wright, D.: Ethics and privacy in ai and big data: implementing responsible research and innovation. *IEEE Secur. Priv.* **16**(3), 26–33 (2018)
55. The Institute for Ethical AI in Education. Interim report: towards a shared vision of ethical AI in education. <https://www.buckingham.ac.uk/wp-content/uploads/2020/02/The-Institute-for-Ethical-AI-in-Educations-Interim-Report-Towards-a-Shared-Vision-of-Ethical-AI-in-Education.pdf>. (2020)
56. Waller, M., Paul, W.: Why predictive algorithms are so risky for public sector bodies. Social Science Research Network, Rochester (2020)
57. Weale, S., Stewart, H.: A-level and GCSE results in England to be based on teacher assessments in U-turn. *The Guardian*, (2020).
58. Zimmerman, M.: Teaching AI: exploring new frontiers for learning. International Society for Technology in Education, Portland (2018)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.