International Journal of Child-Computer Interaction 34 (2022) 100541

Contents lists available at ScienceDirect

International Journal of Child-Computer Interaction

journal homepage: www.elsevier.com/locate/ijcci

Characterizing children's conceptual knowledge and computational practices in a critical machine learning educational program

Golnaz Arastoopour Irgens *, Hazel Vega, Ibrahim Adisa, Cinamon Bailey

Education and Human Development, Clemson University, Clemson, SC, USA

ARTICLE INFO

Article history:
Received 6 May 2022
Received in revised form 8 September 2022
Accepted 20 September 2022
Available online 27 September 2022

Keywords:
Critical pedagogies
Machine learning education
Elementary education
Design-based research
Robotics
Informal education

ABSTRACT

In this study, we describe the design and implementation of a CML (critical machine learning) education program for children between the ages of 9 and 13 at an after-school center. In this participatory design-based research, we collected learner artifacts, recordings of interactions, and pre/post drawings and written responses to model children's developing knowledge and practices related to critical machine learning. Drawing from constructionist and critical pedagogical perspectives, our research questions are: (1) How do children develop machine learning knowledge grounded in social, ethical, and political orientations in a CML education program? and (2) What computational practices do children engage in when developing robots for social good in a CML education program? We found that (1) children made more sophisticated connections with socio-political orientations and ML content as they progressed through the program, and (2) they engaged in computational practices, such as experimenting and iterating, testing and debugging, reusing and remixing, and abstracting and modularizing. Further, our findings indicate that a critical lens to ML education can be characterized by posing and answering questions about the roles of AI technologies producers and consumers and identifying how these technologies are designed to apply this knowledge to build applications for marginalized populations. This study suggests that a critical lens is an effective approach towards engaging young children in designing their own machine learning tools in socially responsible ways.

1. Introduction

Artificial Intelligence (AI) has become increasingly common in digital technologies. Such technologies improve our quality of life by automating tasks and augmenting human capabilities. For example, government and media giants like Google, Facebook, Netflix, and Twitter to collect large amounts of data about people by monitoring and tracking their online activities (Frary, 2017). Powerful machine learning algorithms subsequently use this data to curate personalized news feed for people and make critical decisions about them (Bucher, 2017). However, while AI is efficient for displaying relevant information and predicting people's preferences at any given time, it also risks being used as a tool for government surveillance (Duberry, 2022) and for manipulating the psychological and physiological behavior of people (Kramer, Guillory, & Hancock, 2014). Moreover, when AI applications are used at large scales to determine whether people receive employment, obtain loans, or are convicted of a crime, they can (re)perpetuate social inequities (O'Neil, 2016). For example, facial recognition applications have been used on user's digital photo collections to identify faces and help users search through thousands of photos (Pirrung, et al., 2018). However, serious ethical considerations with facial recognition applications have materialized. Critiques include concerns about privacy and the increasing culture of surveillance (Smith & Miller, 2022). Others are concerned with the racialization of focusing on faces (Stark, 2019) and the reinforcement of existing racial disparities for historically disadvantaged groups (Bacchini & Lorusso, 2019).

These concerns go beyond facial recognition applications and are widespread with those who work with AI. However, the general public does not have a solid understanding of how these technologies work and the potentially harmful effects of large-scale AI deployment (Fox-Skelly, et al., 2020). Children in K-12, particularly at the elementary school level, are not exposed to ethical AI issues even though they use these technologies in their daily lives. Without an education that includes reflection on the social and ethical consequences of AI technologies, future generations will continue in harmful traditions of technology consumption and development without a critical lens.

In this work, we define *critical machine learning* (CML) education as machine learning education that centers social, ethical, and political orientations in AI. This approach to teaching and learning is grounded in critical pedagogy (Freire, 1970; Giroux, 1985) and influenced by constructionist design philosophies

* Correspondence to: College of Education, 101 Gantt Circle, 29634, USA.
@clemsonidealab (G. Arastoopour Irgens).
E-mail address: garasto@clemson.edu (G. Arastoopour Irgens).

https://doi.org/10.1016/j.ijcci.2022.100541

(Papert, 1980; Papert & Harel, 1991). In this study, we describe the design and implementation of a CML education program for children between the ages of 9 and 13 at an after-school center. By collecting learner artifacts, recordings of interactions, and pre/post drawings and written responses, we model children's knowledge and practices related to critical machine learning. The research questions are: (1) How do children develop machine learning knowledge grounded in social, ethical, and political orientations in a CML education program? and (2) What computational practices do children engage in when developing robots for social good in a CML education program?

2. Background and theory

2.1. AI competencies and computational practices for children

Artificial intelligence is a broad term for describing machines that think or act like humans. More precisely, AI systems are software and/or hardware systems designed by humans that perceive their environment through data acquisition and then rely on this data to take actions to meet a given goal. Humans can program AI systems using symbolic rules, train them to use numeric models, or train them to adapt their behavior by analyzing how the environment is affected by their previous actions (The European Commission, 2019). Machine learning (ML) is a practical implementation of AI. Humans build ML models based on training data to make predictions or decisions without being explicitly programmed to do so (Samoili, et al., 2021).

Although AI has been assimilated into everyday technology use, there are no agreed upon K-12 AI and ML learning standards. However, Touretzky and colleagues (Touretzky, Gardner-McCune, Martin, & Seehorn, 2019) have outlined "five big ideas" for K-12 education. Specific to grades 3–5 and 6–8, they suggest that students should be able to modify simple sensor-based applications in children's programming frameworks (3–5), use children's programming languages to create a series of conditional statements or decision trees (3–5), modify object recognition applications and measure how well a trained systems generalizes to novel inputs (6–8), and understand how biases in training data can affect performance (6–8). Drawing on this work, Long and Magerko (2020), developed 17 competencies for AI through an exploratory literature review. These competencies include recognizing technology that does and does not use AI, imagining future applications and effects of AI, and recognizing that humans play an important role in AI development.

Additionally, scholars have argued for children's engagement with computational practices and have developed tools specifically for this purpose (Grover & Pea, 2013; Kafai, 2016). For example, Scratch is a block-based programming tool in which users drag and drop puzzle pieces onto a workspace to created executable programs. Instead of using text to write algorithms, children rely on the constraints of how the blocks fit together as clues to guide their programming. Empirical work with Scratch learners (Brennan & Resnick, 2012) suggests that children engage in at least four sets of computational practices: experimenting and iterating, testing and debugging, reusing and remixing, and abstracting and modularizing (Fig. 1).

Extending Scratch into AI and ML education, findings from Shamir and Levin (2022) showed that elementary school students engaged in computational practices by creating ML classifications systems. Including ethical components in ML education, Williams and colleagues (Williams, 2021) developed Scratch blocks specific for AI topic exploration and designed a corresponding curriculum. The curriculum contained activities for children in 5th–8th grade to explore AI, ML, and ethics through robotics.

2.2. Taking a critical pedagogy lens to machine learning education

The current body of theoretical and empirical work is bringing ethics into the conversation for computer science and AI education. However, ethical and humanistic orientations are not fully integrated throughout the proposed competencies nor tools and curricula. This isolation of ethics is aligned with Borenstein and Howard's (Borenstein & Howard, 2021) critique that "ethics should not be a slapped-on component after-the-fact, a standalone lesson, or a second thought. It is integral at every stage when learning about AI" (p.62). It is integral at every stage because developing AI applications is fundamentally a human and social endeavor. Every AI application was developed by a person with a particular worldview that influenced their decision-making during development. That person's assumptions, opinions, and biases are embedded into the tool (Pea, 1993). Importantly, those involved in designing and developing AI systems are predominantly White and Asian men with beliefs and values that may not necessarily represent that of a diverse society (Raub, 2018). These individuals often end up embedding their opinions into algorithms deployed and used to make critical decisions about the public. When human designers are not made visible, then such tools may incorrectly appear singularly truthful and free of bias to the users (D'Ignazio & Klein, 2020). Haraway (1988) critiques this removal of bodies from knowledge as "a view from nowhere". As such, dismembering AI applications from the bodies that produce such technologies provides an incomplete view of AI and thus, does a disservice in terms of educating our youth.

In addition, the ethics explorations in computer science education tend to focus on a limited form of "microethics", centered on individuals making decisions when faced with dilemmas (Vakil & Ayers, 2019). This narrow approach ignores the sociopolitical contexts of how technologies are developed and presents technologies as ahistorical and neutral. However, all systems, including those in computer science, are embedded in existing politicized social systems (D'Ignazio & Klein, 2020). For example, Buolamwini and Gebru (2018) evaluated three commercial image classification systems used for facial recognition technology. The study was spurred by Buolamwini's personal experiences as a Black woman being misidentified when using facial recognition software. The researchers found that darker-skinned females were the most misclassified group with error rates up to 34%, while the maximum error rate for lighter-skinned males was 0.8%. These error rates become particularly concerning when facial recognition systems are being used by U.S. government agencies to detect unlawful behaviors and thus, (re)enforcing historic inequities in law enforcement against people of color.

To make social, ethical, and political issues central to the design of ML educational environments, we draw from critical pedagogy theorists who argue that teaching and learning are inherently rooted in social, historical, political, and economic contexts (Freire, 1970; Giroux, 1985; Vakil & Ayers, 2019). The dominant worldviews, such as those from White men and the middle-class, tend to be adopted in educational institutions. In contrast, perspectives from marginalized populations, such as people with histories of slavery, with histories of colonization, who live in poverty, and people of color, are less visible. Because of this power differential, schools may reproduce inequities and provide fewer opportunities for the oppressed. From a critical pedagogy view, educators and learners must "wake up" to become aware of the oppression happening to and around them and co-create new knowledge together. Aligning with this perspective, we propose a *critical machine learning* (CML) educational approach that integrates critical pedagogy into computer science and machine learning education. In this integrated approach, machine learning knowledge and computational practices are not

experimenting and iterating developing a little bit, then trying it out, then developing more	**testing and debugging** making sure things work – and finding and solving problems when they arise
reusing and remixing making something by building on existing projects or ideas	**abstracting and modularizing** exploring connections between the whole and the parts

Fig. 1. Definitions of four computational practices from learners using Scratch.
Source: Image recreated from https://scratched.gse.harvard.edu/ct/defining.html.

separated from the social, historical, and political contexts in which people consume and produce technology. Rather, when learners apply a critical pedagogy lens to AI technologies, they focus on questions such as, Who develops these technologies? What are the developers' interests? For whom are these technologies designed? What types of data are used to train machines? What is the history behind the data used? What decisions are made based on the outputs of the algorithms? Educators and learners pose such critical questions, reflect, discuss, and co-develop solutions to disrupt oppressive paradigms related to the development and consumption of modern digital technologies.

2.3. Constructionism

The CML approach relies on constructionism to guide the design of the learning environment. Constructionism emphasizes creating "objects-to-think-with" that represent how a learner actively (re)constructs their understanding of a domain (Kafai & Resnick, 1996; Papert & Harel, 1991). In most cases, the object that is being constructed is computational in nature (Holbert & Wilensky, 2019; Wilensky & Reisman, 2006) and can be manipulated in multiple ways to represent conceptual ideas (Papert, 1980). Additionally, when learners have access to multiple representations of concepts, they can make decisions about how to connect among these representations and pieces of their knowledge. The more connections learners make between objects, the richer their understanding of the underlying concepts related to that object, and ultimately, the higher the quality of the relationship with the object and concepts (Wilensky, 1991).

Constructionism aligns well with a CML educational approach because learners can create their own computational ML objects-to-think-with that "support counter-narratives to existing dominant ideologies" (Lee & Soep, 2016) and pose questions about the social, ethical, and political nature of such objects. When creating their own ML tools, learners can tinker and receive feedback from the tool to develop their computational skills. As learners are given the space to (re)construct their objects and their understanding of the underlying concepts, they will develop personal relationships with the ML concepts and ideally, fuse their own interests and histories into the objects. Although not directly focusing on AI and ML, scholars have investigated the intersections of constructionism, computational literacies, and critical pedagogies (Blikstein, 2008; Blikstein & Blikstein, 2021; Lee & Soep, 2016). For example, Vakil (2014) discovered that engaging youth in exploring the sociopolitical contexts of computing, such as designing mobile apps to address issues in local communities, can lead to deeper emotional engagement. Furthermore, he claimed that a critical pedagogy lens applied to developing digital technologies is a powerful and flexible method for eliciting multiple pathways for students to move beyond emotional engagement and into cognitive engagement.

Our approach of CML education differs from existing critical CS/data education frameworks such as critical computational expression (Lee, Gobir, Gurn, & Soep, 2022; Lee & Soep, 2016), critical algorithmic literacies (Dasgupta & Hill, 2021), and critical data literacies (Hautea, Dasgupta, & Hill, 2017; Stornaiuolo, 2020) because the focus is specifically on understanding how *elementary-school aged children* engage in *AI/ML applications* that rely on data *curated by humans* and the *sociopolitical contexts* of the deployment of such applications. As noted above, such AI applications have ethical and sociopolitical consequences for non-dominant populations. Meanwhile, there is little research on how children examine and critically engage with sociopolitical issues in computer science knowledge and practices within the context of AI/ML applications. Although CML education has not yet been implemented with elementary school-aged students, studies suggest that children have the ability to engage with sociopolitical issues. Starting at infancy, children notice differences in terms of race and gender, and by early elementary school, children judge ingroup members who look like them more favorably than outgroup members (Dunham, Baron, & Carey, 2011). Classroom intervention research suggests that upper elementary school aged students are able to consider oppression from multiple perspectives including the broader historical framework of how society is organized and how to create change (Fain, 2008), analyze and interrogate literature around societal issues such as immigration (Braden, 2019), and address and challenge social inequities in their own curricula (Kersten, 2006).

3. Methods

3.1. Participants

We implemented this research project in after-school care programs at two separate community centers that serve elementary schools in a Southern U.S. county with a mix of urban and rural areas, a poverty rate of 13.4%, a household median income of $56,609, and a population that is 67% White (non-Hispanic), 23% Black, 5% Hispanic, and 2% Asian. Participants between the centers included 44 youth, 3 staff counselors (2 in one center, 1 in the other), and 4 researchers (2 at each center). Not all youth participated in every activity session due to other obligations, such as homework or after-school sports programs. The youth population at both centers consisted of Black, Latinx, and White children with a mix of those who presented as girls and boys and ranged between 9–13 years of age, with about 80% between ages 10 and 11. All children and center names are pseudonyms. Researchers were university faculty and graduate students and consisted of a White/Middle Eastern woman, White woman from the local region, Nigerian Black man, and Costa Rican Latina woman. The lead author and director of the research project is a former computer science and mathematics instructor whose perspective and passion has influenced the design of the current program. The study was conducted at both centers for seven weeks and consisted of pre-post drawings about algorithms and ML; 9 activity sessions held on separate days, each lasting approximately 2 h; and post-interviews with the participating children. This was the third iteration of an ongoing participatory design study.

3.2. Research design and context

This study adopts a design-based research method that generates and tests learning theories in natural contexts (Lee & Soep, 2016; Wilensky, 1991). In addition to advancing theory, design-based approaches directly impact practice and social change (Blikstein, 2008). In this study, the anticipated impacts were to co-develop and disseminate a CML educational program and provide children with an opportunity to gain critical computer science knowledge and practices. The study occurred in an informal learning context in which researchers encouraged youth to learn actively through production and discovery. Through incremental activities developed by researchers, youth engaged with ideas around how machine learning systems and structures create and sustain societal inequities. For example, youth created a facial recognition machine for cats and dogs and used training datasets that were highly favorable for cats but biased against dogs. They were then encouraged to explore why the training datasets were biased, what the consequences of the deployment of such biased algorithms would be, and how to mitigate the bias in future designs. These ideas were then extended to humans, when youth watched a video about the film, Coded Bias (Kantayya, 2020), in which Black citizens were discriminated against with facial recognition software and experienced negative consequences in their lives. We anticipated activities such as these would foster the development of critical consciousness: the ability to recognize and analyze systems of inequality and engage in social change making, thinking, and behavior (Freire, 1970) in the context of technological systems (Blikstein, 2008).

To develop the CML education program, we adapted activities from MIT's How to Train Your Robot (https://www.media.mit.edu/projects/ai-5-8/overview/) and AI+Ethics for Middle School Curriculum (https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/). We met with the designers of both curricula to ensure our values and motivations around guiding children to develop critical lens with technologies were aligned. We also relied on prior participatory design activities with children in 2020 before COVID-19 pandemic restrictions began (Blikstein & Blikstein, 2021). Children were engaged in and contributed to the participatory design process as users, testers, and informants (Druin, 2002; Guha, Druin, & Fails, 2013). Children were observed and questioned as they experienced the program, and this data was used to redesign the subsequent activities and then the program as a whole during and for the next iteration. The research team met weekly to review the discussions, activities, and children's interactions that took place during the previous session and planned the following week's session. We jointly reviewed the transcripts and videos of each session to study the children's thinking, levels of engagement, and learning interests. The program began in one center two weeks prior to the other center. This staggered implementation allowed us to implement changes across centers when problems arose or when data suggested we do so. For example, after the Interest Board Activity in the first center, researchers noted the children (ironically) did not appear interested and did not list interests that could be implemented into the CML program, which was the goal of the activity. Thus, this activity was not utilized in the second center. Rather, researchers spent additional time playing with the children with robots in order to understand their existing ML knowledge and discuss their interests while building relationships. Post-interviews with the children provided opportunities to further gather input regarding their perceptions on what they did and did not enjoy during the program and what they learned throughout the program. Using these participatory design and cooperative inquiry techniques, we made redesign decisions both during and after implementation that reflected the voice, preferences, and interests of the target end users of the CML program, while also negotiating our own research goals.

Table 1 summarizes the CML activities that children engaged in during the study. The pre- and post-drawings were designed to assess learners' knowledge of algorithms and machine learning before and after the implementation of the CML activities. To complete the drawings, children used drawing tools such as crayons, markers, and/or pencils. In the interest boards activity, children showcased their interests by creating handmade cut-out pieces of pictures from magazines or books and gluing them to construction paper. In the pizza algorithm activity, children were asked to write an algorithm to make the best pizza using posters and markers. This created an opportunity for the children to explore what it means to be the "best" and see how their opinions were reflected in their algorithms. In the helpful and harmful technology activity, children listed and drew everyday digital technologies on posters and reflected on whether these technologies are harmful or helpful and why. During the Google Search activity, children searched different topics using the Google search engine and discussed representation and bias issues related to these searches related to gender and race. The Google's Quick Draw activity introduced children to the concepts of training data, input, and output by experimenting with a game that guesses a user's drawing (https://quickdraw.withgoogle.com/). In the cat and dog activity, children built a cat-dog classifier using Google's Teachable Machine (https://teachablemachine.withgoogle.com/) but were unknowingly given a biased dataset. When the classifier worked more accurately for cats than dogs, children retrained their classifiers with new datasets that were less biased. This activity was designed to help children understand that machine learning is dependent on the training data that is being used which in turn, determines the effectiveness of the algorithm. In this case, the biased training dataset misclassified and excluded certain breeds of dogs. For the build-your-own teachable machine activity, children created a classification machine that recognized images, poses, or sounds using Google's Teachable Machine. Children trained their machines using items or images of their choice. They also tested their peers' machines for functionality and bias. Next, children watched the Coded Bias film trailer, which features Joy Buolamwini's realization of racist facial recognition technologies [52]. Children, staff, and researchers engaged in a large group discussion afterwards regarding representation and bias in AI facial recognition systems. Last, using markers and posters, children created narrative stories about robots that can be helpful to people, which we referred to as "superhero" robots. They then created a narrative about their robot's superpowers. To build their design, children experimented with micro:bit robots using Scratch block-based programming software to program their robot. Table 1 summarizes the CML activities. The design of the CML educational program is described in more detail in Arastoopour Irgens et al. (2022) and the full program description can be accessed at idealabclemson.org.

3.3. Data collection and analysis

To answer RQ1: How do children develop machine learning knowledge grounded in social, ethical, and political orientations in a CML education program?, we collected children's written

Table 1
CML educational program activities and descriptions.

Program activity	Activity description
Pre and Post Drawing and Response	This is an independent task that required sketching to access learners' understanding of ML knowledge and issues and asked them to respond to two short answer questions.
Interest Board	These were handmade creations by the youth illustrating what they were interested in.
Pizza Algorithm	Youth work in groups to write an algorithm to make the best pizza. This helps to explore what it means to be the "best" and see how their opinions are reflected in their algorithms.
Harmful/Helpful Technologies	In this activity, youth discuss the technologies they use or see in their everyday lives. They are asked to reflect on whether these technologies are harmful and helpful and why.
Google Search	Youth are guided to search different topics using Google search engine and discuss representation and bias issues related to these searches.
Google's "Quick Draw!" Activity	Youth are introduced to the concepts of training data, input, and output. Focus is on direct interaction of humans with machines in machine learning and algorithmic processes.
Cat and Dog Teachable Machine	Youth are introduced to the concept of classification; they build a cat-dog classifier but are unknowingly given a biased dataset. Youth investigate how based training data are unfair.
Build your own Teachable Machine	Youth create their own machine that recognizes images, poses, or sounds using Google's Teachable Machine. Youth train their machines using items or images that they choose. They also test their peers' machines for functionality and bias.
Watch and Discuss Coded Bias Trailer	Youth watch a trailer to further learn what bias is and how some facial recognition technologies/AI technologies are biased keywords
Robot superhero stories	Youth develop stories of how a "superhero" robot can be trained to complete a task and help people.
Machine learning robot for social good	Youth create their prototype robot based on their robot stories that recognizes images, poses, or sounds and try to win a prize.

responses and drawings from all activities. Because the integration of critical pedagogy and computer science education is still undertheorized, children's responses were coded using a social constructionist grounded analysis (Charmaz, 2008) with CML motivations as a guiding framework. The data were segmented by each child's response. After segmentation, we engaged in several iterations of emergent coding, explicating analytic and methodological decisions. One researcher coded the dataset, and another reviewed the coded data. The coded data were quantified; if a code existed in a response, it received a "1" in that code category and if a code did not exist in a response, it received a "0" in that code category. Nine activities had responses that appeared in the codes: the 3 pre drawing/short answer activities, Google Quick Draw, Cat and Dog Teachable Machine, robot superhero stories, and 3 post drawing/short answer activities.

After quantifying the coded data, we used the Epistemic Network Analysis (ENA) 2.0 webtool (Marquart, Hinojosa, Swiecki, & Shaffer, 2018) to measure and visualize the connections children made among CML conceptual knowledge codes in their discourse. ENA measures the connections between discourse elements, or codes, by quantifying the co-occurrence of those elements within a defined stanza (Shaffer, 2017, 2018). Stanzas are collections of utterances that are topically related. Once the size of a stanza is identified, for any two codes, their strength of association is computed based on the frequency of their co-occurrence within each stanza in the discourse. In this study, we defined a stanza as one response from a child. Thus, co-occurrences of codes were calculated if they occurred within a child's response. After defining the stanza, each child's co-occurrences for each of the nine activities were summed and each activity was visualized as a weighted node-link network representation. This single network represented a summation of all the children's co-occurrences within an activity. To analyze several networks at one time, we used an alternative ENA representation in which the centroid (center of mass) of each network was calculated and plotted in a fixed two-dimensional space that was mathematically created by conducting a multi-dimensional scaling routine and a sphere-normalization. The space is interpreted by examining the location of the nodes in the two-dimensional space and evaluating the goodness of fit. In this analysis, the Spearman goodness of fit was 0 for both the x and y axis and the Pearson goodness of fit was 1.0 for both the x and y axis, indicating the location placement of the nodes was reliable. For more detailed mathematical explanations of ENA, see work by Bowman and colleagues (Bowman, et al., 2021), Arastoopour Irgens and colleagues (2021), and Shaffer and Ruis (2017).

To answer RQ2: What computational practices do children engage in when developing robots for social good in this CML education program?, we collected video recordings of children programming their robots. We focused on one video of two fourth grade girls for the methodological reason of providing a detailed, microgenetic discourse analysis of practices which complemented the broader analysis of conceptual knowledge addressed in RQ1. This video was also chosen for pragmatic reasons as it was the only audible video evidence collected of children's computational practices recorded from children who assented to being videotaped and whose parents consented. To analyze the video, we relied on Discourse analysis (Gee, 2011), which assumes that language is situated and social. Gee (2008) argues that people make sense of the world by engaging in big "D" Discourses which are combinations of language, actions, ways of thinking, ways of being, valuing, and using tools. People use these forms of Discourses in different ways for different purposes. Children, in particular, may use a variety of "everyday"

Discourses adapted from their home, school, and community lives which shape their perspective and values and can be leveraged for learning (Rosebery, Ogonowski, DiSchino, & Warren, 2010). In this study, we transcribed the audible language used in the video and segmented the transcript by turns of talk. We used the transcription as a reference tool when analyzing the video. During the video analysis, we noticed intonations, gestures, eye gaze, and physical positions. Integrating these forms of Discourses, we described how children engaged in four sets of computational practices: experimenting and iterating, testing and debugging, reusing and remixing, and abstracting and modularizing (Brennan & Resnick, 2012).

4. Findings

4.1. Integrating machine learning knowledge with social, ethical, and political orientations

In this section, we address RQ1: How do children develop machine learning knowledge grounded in social, ethical, and political orientations in this CML education program?

After several iterations of coding and conceptualizing, we finalized a CML conceptual knowledge coding scheme consisting of three macro-categories of codes: How people develop machine learning applications (5 codes), harmful machine learning applications (3 codes), and helpful machine learning applications (2 codes) (Table 2).

The networks illustrate the patterns of discourse that occurred in each activity. In the pre activities there were no connections made among the codes within any children's written responses and thus, there is no discourse network pictured. In the following activity, Quick Draw, children explored an online application that guessed their drawings. The children connected between the codes: ALGORITHM LIMITATIONS and DIVERSE USERS (Fig. 2), suggesting an understanding of how the performance limitations of ML algorithms could exclude certain people. For example, in the Quick Draw activity, when asked "Will this machine work for everyone?" Jasmine wrote, "People draw differently, and the algorithm can't notice everything". In her written response, Jasmine explained that the algorithm did not have the ability to guess every possible object that a person may draw because of the limited training dataset that the developer used to train the machine. Those people that "draw differently" than what the algorithm recognized, would not be able to use Quick Draw. In this activity, children connected across two categories coding categories: Harmful Machine Learning Applications (nodes labels are red) and How People Develop Machine Learning Applications (node labels are grey), suggesting a connected understanding of how people can train machine learning algorithms that could exclude certain populations. These connections reflect the goal of the designed activity for children to tinker with and critique an existing machine learning based application.

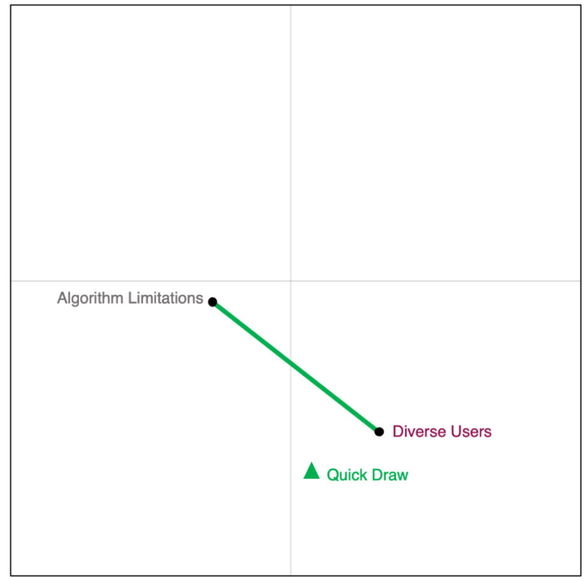


Fig. 2. Discourse network for Quick Draw activity. Triangle represents the center of mass of the weighted network and is plotted in a two-dimensional mathematical space. Node labels for Harmful ML Applications codes are red and for How People Develop ML Applications are grey. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In the next activity, Cat Dog Teachable Machine, children trained a machine to classify images of cats and dogs. The training dataset was purposefully biased against dogs and recognized images of cats more accurately than dogs. After training and testing their machines, children answered the question, "Why did this machine work better for cats than dogs?" In their discourse networks, children connected TRAINING DATA BIAS BY VARIETY to TRAINING DATA BIAS BY NUMBERS and ALGORITHM LIMITATIONS (Fig. 3), suggesting an understanding of how training data can be biased in at least two ways and these biases contribute to the limitation of ML algorithms. These connections reflect the goal of the designed activity for children to create and tinker with a machine learning based application and reflect on its limitations in terms of biased training data. For example, Amber wrote, "Cats look very different, and computer doesn't have human intelligence". Here, Amber provided two points of reasoning for why the machine worked better for cats than dogs. She identified a limitation of ML algorithms by claiming that although computers are developed by people, they do not "have human intelligence" and thus, machines do not classify cats and dogs as effectively as humans, or at least in the ways that humans perceive as effective. Amber also claimed that the training dataset contained a variety of cats that "look very different", meaning that Amber provided the machine with a wider range of images to learn from when classifying cats compared to dogs. In another example, Lisa noticed that in addition to variety, the cat classification training dataset had a higher number of cat photos. She wrote, "More pictures of cats and more kinds of cats".

In the third activity, Robot Superhero Story, children created and illustrated their own stories about a superhero robot that could help people. In this activity, they connected among TECHNOLOGY FOR KIDS, TECHNOLOGY TO ADDRESS SOCIAL ISSUES, CLASSIFICATION ALGORITHMS, and DIVERSE USERS suggesting an understanding of how machine learning classification algorithms can be used to design technology for kids, for social good, and for a diverse range of users (Fig. 4). These connections reflect the goal of the designed activity for children to imagine and design their own machine learning based robot that could benefit a particular, potentially marginalized, population. For example, LaToya, an African American child, designed a robot that relied on a color classification algorithm that would identify the colors of real-world objects and teach colors to young children. She was inspired by her younger cousin's lack of access to learning tools. In a conversation with a researcher, she explained, "My cousin grew up with my grandparents because her mom died a couple of years ago. When she was growing up, she didn't have the opportunity to, like sit down every day and watch TV shows that teach her colors and stuff. And so, the only time she had stuff to learn is when I came down with my books and like taught her. And so I thought to myself, that could be happening to multiple other kids all over America. And so, I thought, well, maybe I

Table 2
Final coding scheme of CML conceptual knowledge with three categories.

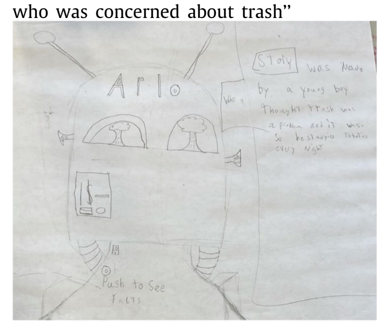
Category	Code	Definition	Written example
How people develop machine learning applications	Algorithm Limitations	AI technology cannot do everything that humans can do, and AI technology cannot act like we intend	"because it doesn't know how to draw"
	Algorithms are made by humans	People program AI technologies. AI algorithms have people's biases, assumptions, and opinions embedded in them.	"because people have different opinions when it comes to coding a robot or a facial recognition site."
	Training data bias by numbers	Training dataset can be biased towards a group if there are not enough data points provided in the training dataset from that particular group. Or if a training dataset has a higher number of datapoints from one particular group, then that group has an advantage.	"there are more cat pictures"
	Training data bias by variety	Training dataset can be biased towards a group if there is not enough variety of datapoints provided for that particular group. Or if a training dataset has more variety from one particular group, then that group has an advantage.	"more pictures of cats and more kinds of cats"
	Classification Algorithms	A subset of machine learning algorithms classifies objects and, if trained, are able to interact with the world to identify classes.	"Face recognition" 
Harmful Machine Learning Applications	Diverse users	AI technology does not always work as intended for all users. Users of technology can come from different cultures, backgrounds, and experiences and this can mean a different experience with the technology. Different than what was intended by the designer of the technology.	"Someone from a different country might draw something differently or you can be bad at drawing"
	Racial discrimination	AI technology may discriminate against people based on their race or ethnicity. Historically, people of color have been marginalized or harmed.	"Like in c0de bias it doesn't recognise blacks."
	Gender discrimination	AI technology may discriminate against people based on their gender identity. Historically, women and non-binary people have been marginalized or harmed.	"The computer mostly recognizes the white men. It is not fair for other races and genders like the women in the video."
Helpful Machine Learning Applications	Technology for kids	An idea for AI technology that serves children.	"headphones that will help autistic kids"
	Technology to address social issues	An idea for AI technology that addresses a social issue, locally or globally.	"Arlo the robot picks up trash and recycles it and gives you facts, created by a boy who was concerned about trash" 

Fig. 3. Discourse network for Cat Dog Teachable Machine activity. Triangle represents the center of mass of the weighted network and is plotted in a two-dimensional mathematical space. Node labels for How People Develop ML Applications are grey. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

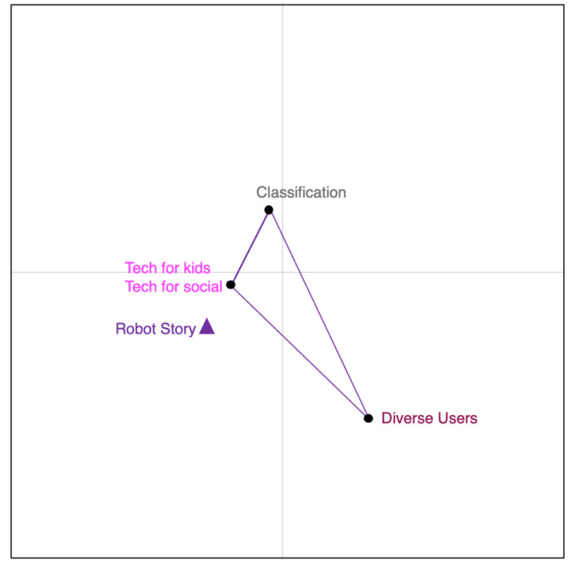


Fig. 4. Discourse network for Robot Superhero Story activity. Triangle represents the center of mass of the weighted network and is plotted in a two-dimensional mathematical space. Node labels for Helpful ML Applications are pink, for Harmful ML Applications codes are red, and for How People Develop ML Applications are grey. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

could make a machine that can help kids with that". Similarly, Ian and Eduardo's robot was designed to help children. Their robot used facial recognition algorithms to classify adults and children and then made decisions based on the results from a diverse set of users. They wrote, "It can be used at home and parents can program it to set the things that it can get to children, they can also set the schedule. For example, it can be programmed to get cookies for kids only between 3 and 5 pm. It needs a face recognition software to identify children and adults". The other children's superhero stories included robots that do homework, clean your room, build homes for those in need, stop COVID-19, provide headphones to help autistic children, and pick up trash for recycling. Overall, in their networks, children connected across all three coding categories of Helpful Machine Learning Applications (node labels are pink), Harmful Machine Learning Applications (nodes labels are red), and How People Develop Machine Learning Applications (node labels are grey), suggesting an understanding that people develop machine learning applications that can be both helpful and harmful.

The final set of activities were the post drawing and short answer responses. The mean network and point for these post activities illustrate connections across several of the codes in two categories (Fig. 5), whereas the pre activities did not reveal any connections. The post network reveals the most connections across all concepts, suggesting a more integrated understanding of critical machine learning by the end of the program. Specifically, children made connections to GENDER DISCRIMINATION, RACIAL DISCRIMINATION, and ALGORITHM BY HUMANS for the first time. These nodes appeared after the children engaged in the Coded Bias film discussion, in which children discussed how machine learning based technology designed by a dominant population can be discriminatory against marginalized populations. The network also reveals that children connected to other nodes that they connected to in previous activities. Specifically, children explained that CLASSIFICATION algorithms can be created for a set of DIVERSE USERS but can also results in RACIAL DISCRIMINATION and GENDER DISCRIMINATION because they are ALGORITHMS CREATED BY HUMANS and have ALGORITHM LIMITATIONS.

For example, when asked to draw a picture of machine learning, Justin drew a picture that resembled the Cat Dog Teachable Machine activity (Fig. 6). His drawing contained five images of dogs and three images of cats grouped as training datasets. In between the two classes, Justin drew two bars that resembled the classification confidence percentages from Teachable Machine, labeled "cats" and "dogs". The bar for dogs was fully colored in, suggesting 100% confidence in classification, whereas the bar for cats was partially colored. Justin's drawing represents a connection between understanding how machine learning can be used to classify images and understanding that there are limitations regarding the accuracy of such algorithms. Similarly, in the post short answer responses, children discussed machine learning classification algorithms but connected this concept to racial and gender discrimination. For example, when asked to provide examples of unfair algorithms and who they help/harm, Lisa referenced the *Coded Bias* film trailer and wrote, "In the video we watched, the girl realized that the computer didn't recognize her face until she put her white mask on. It wasn't fair because most of the pictures of faces were white men. The computer mostly recognizes the white men. It is not fair for other races and genders like the women in the video". Lisa identified facial classification as a machine learning algorithm that could be unfair to certain populations. She claimed that "white men" benefit in this scenario because they are recognized by the algorithm and that "it is not fair for other races and genders" who are not recognized by the algorithm. Similarly, LaToya referenced the film trailer and recognized how Black bodies were excluded from some machine learning applications. However, LaToya extended her understanding by discussing the harmful social consequences of racial misclassification. She wrote, "it can deny people property and housing and jobs and can really effect and change people's lives due to the fact that they were different to the person that made the algorithm". In her response, LaToya addressed that people create machine learning algorithms, and if the people who make them are "different" than those who use them, the consequences can "really affect people's lives" in harmful ways by denying them "housing and jobs". In other words, she expressed

Fig. 5. Mean discourse network of children's post drawings and short answer responses.

how people in positions of power can design machine learning algorithms that misclassify those who are not in positions to design them. Those misclassifications can lead to further disadvantaging underprivileged populations. Overall, in this final set of activities, children connected across several codes within the categories of Harmful Machine Learning Applications (nodes labels are red) and How People Develop Machine Learning Applications (node labels are grey), suggesting a thicker understanding of how people train machine learning algorithms that could exclude or harm others.

4.2. Computational practices in the context of designing AI for social good

In this section, we address RQ 2: What computational practices do children engage in when developing robots for social good in this CML education program? by examining how Madison, Amber, Emma, and Lisa designed a robot for social good and how Madison and Amber created and tested a prototype of Ted using the micro:bit robot and Scratch AI programming blocks. To connect this discourse analysis of computational practices with the analysis of machine learning knowledge above, we highlight the knowledge codes that children connected to in their practices.

4.2.1. Designing ted, the helpful robot

When creating their robot story, the girls imagined a helper robot that they named Ted (Fig. 7), who was a human-size robot controlled by voice activation. For example, when the robot heard "Hey Ted", he responded by saying, "Hi, what can I do for you?" Ted could complete almost an endless number of tasks. When a researcher, Ophelia, approached the group, the girls collectively explained, "He makes our breakfast, he wakes us up, he brushes our teeth, he puts groceries away, he babysits when your parents are gone, do your homework, he can set timers and alarms, he can do the laundry for you. We just came up with the name off the top of our heads".

After describing Ted, they talked about the limitations of their robot. Madison explained, "He can't get wet, so you have to put clothes on him". Emma gave an example of the types of mistakes that Ted may make during childcare. She explained that you can ask Ted "watch my baby for so and so hours, and if you don't get home, he just leaves your baby". Madison was appalled and exclaimed, "No, he doesn't!" Ophelia, the researcher, laughed and stated that it was a good thing to think about the limitations of your robot. Here, the children and researcher discussed the knowledge code of Algorithm Limitations when imagining and designing an AI robot for social good. They also drew on knowledge about creating Technology for kids and Technology for addressing social issues, when they explained how Ted could help children with their daily tasks.

4.2.2. Initial tinkering

When the girls were invited to program a prototype of Ted, they split into two pairs. This analysis focuses on one pair: Madison and Amber. The girls shared one laptop, and Madison programmed, while Amber sat next to her (Fig. 8). After tinkering with the robot to experiment with the different functionalities, the girls decided to focus on one helpful aspect of Ted: that he could empathize with people's feelings and "cheer people up if they were sad". To accomplish this goal, the girls explored the blocks, searching for a way to train a machine learning algorithm that could enable Ted to hear them and respond. Using their prior knowledge about Scratch, they used a sensing block that accessed the laptop's microphone to detect sound. However, the block was not specifically designed for programming the robot. Thus, when Madison tested the sensing block by screaming "Ted!" loudly, she became frustrated and softly stated to Amber, "It's not working". At that moment, Ophelia, a researcher, walked by with her smart phone in her hand and repeated to the girls, "It's not working?" She stopped and offered to help the group.

4.2.3. Ted talks

Madison showed Ophelia their code, which had 10 robot function blocks, such as changing the headlight colors or moving forward for 3 s. Ophelia reviewed the code, pointed to the audio sensing block, and said, "I don't know what that block does. You want the robot to hear you, right? I'm not sure how to get the robot to hear you. I'm looking it up now". Madison and Amber tinkered with the blocks while Ophelia searched for a solution on her smart phone. After a full minute, Ophelia explained that there was a series of blocks that would respond to an audio input. By pointing at the screen, she guided Madison towards removing some blocks and adding others. Together, they created a simple voice recognition algorithm in which the robot first asked, "How are you?", waited for a response, displayed the response on the screen, and then repeated the response. Ophelia tested this algorithm by responding to Ted and said "I'm good. How are you?" When Ted repeated her response, Ophelia shouted, "I'm still good but stop copying me!" The girls giggled, and Madison exclaimed, "That's cool!" Amber scrunched her eyebrows together in skepticism and asked, "Wait, so when we say hello, he will say hello back?" Ophelia did not respond directly to Amber and stated, "Okay you can play around with that for a little bit", and left to assist another group who was calling for her. Here, Amber was using her every day, anthropomorphized language of the robot "saying hello back" to indicate her desire to build in AI components in which Ted would interact with different human inputs.

4.2.4. "Ted, stop copying me!"

After Ophelia left, Madison and Amber tested their algorithm three times. After the third test, Amber softly stated, "I wish it didn't talk..." Sensing Amber's frustration that Ted could only ask "How are you?" and repeat what it heard, Madison replied, "It's because it doesn't have anything else to say. But what if we put this..." Madison dragged a block over to their algorithm which programmed Ted to move forward after listening for audio. She exclaimed, "Amber, I think we are on to something!" Then,

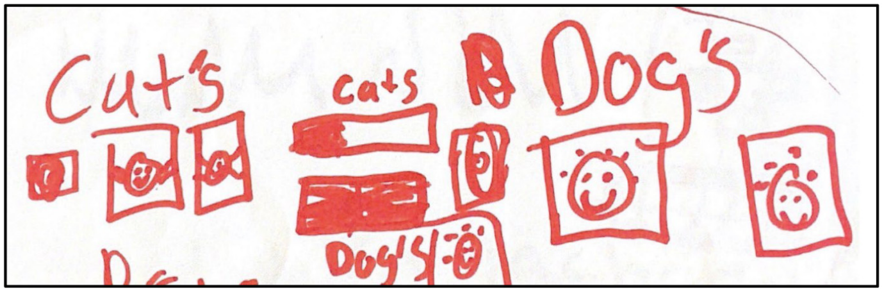


Fig. 6. Justin's drawing of an example of machine learning on a computer. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

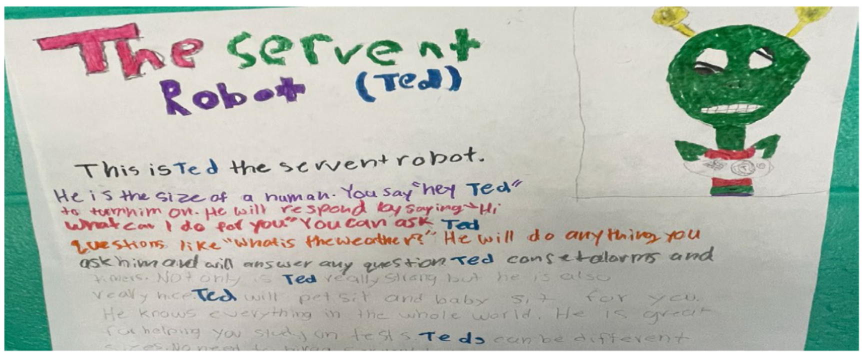


Fig. 7. Madison, Jasmine, Emma, and Lisa's robot superhero story about Ted.

Fig. 8. Madison and Amber working together to program Ted using a micro:bit robot and Scratch AI blocks.

Madison changed the text on one block to change Ted's speech (Fig. 9a). As she made these changes, she communicated to Amber, "So, now it's not going to ask, 'How are you?' It's going to ask, 'Why are you mad?'" When the girls executed their algorithm, Ted asked "Why are you mad?" Madison responded, "Because I am!" Ted repeated, "Because I am!"

Although the girls were successful in changing Ted's initial speech, they were frustrated with how Ted copied their responses. Then, Madison had an idea to remove the "answer" block from within the "speak" block so that Ted would not repeat the "answer" it heard. She stated, "Oh, I know. Be like, 'I'm here to help you".' In place of the answer block, Madison typed "I am here to help you" (Fig. 9b). Amber looked over Madison's shoulder at the algorithm and added, "Oh yeah, because it said that right after", indicating she understood the sequence of the ask-and-listen algorithm. Thus, working on their own, Madison and Amber succeeded in achieving their immediate goals of changing Ted's initial speech, stopping Ted from repeating their responses, and changing Ted's response to something they desired. However, at this point, they changed their algorithm from a simple AI algorithm that responded to the outside environment to a non-AI algorithm that executed the same program regardless of the outside environment.

4.2.5. Integrating conditionals and speech classification models

At that moment, Ophelia walked by again and the girls demonstrated their progress. In the interest of integrating AI into their robot, Ophelia proposed an idea in which Ted could listen and then respond based on what the speaker said. She asked, "What if your robot actually listened to you and if it heard you say, "I'm sad" it would do one thing, but if it heard you say, "I'm happy" it would do something else?" Ophelia then demonstrated what she noticed about Ted, "Yeah, so what's happening here is it's not really listening to you is it? Because I could say whatever I want". Ophelia asked Madison to execute the program. Ted stated, "Why are you mad?" Ophelia responded, "I'm not mad. I'm happy!" Ted responded, "I am here to help you". Ophelia yelled at Ted, "I don't need help!" Then, she turned to the girls, "Right? It's not

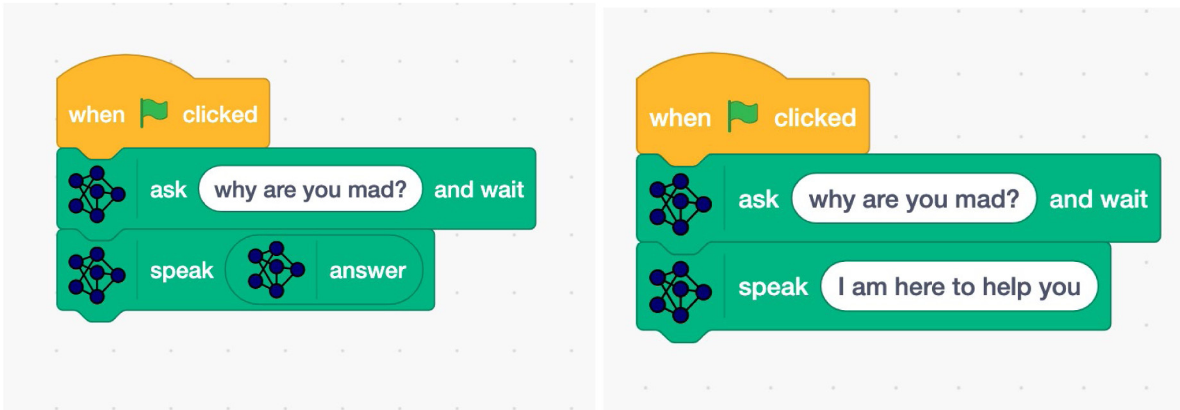


Fig. 9. Madison and Amber's code Scratch AI progression: a) an ask-and-listen algorithm that responds with the "answer" from the user, repeating what the user says, b) a hard-coded ask-and-listen algorithm that repeats the same response each time.

even listening to me no matter what it's going to do the same program, right?" Madison asked, "But how do we know what to program?" Ophelia suggested that they change Ted's inquiry to "Are you happy or sad?" to limit the user's response. She explained, "Because if you ask, Hey, how you feeling? They can say anything, right? But if you ask an OR question, right, then you can control what the person's going to say… So, you're going to need an 'if statement.'" Amber immediately asked, "What's an if statement?" Ophelia answered, "It's like, if somebody says this, then do this". Amber replied, "Oh yeah!" turned to the computer, and after a few seconds, found an if-else statement. She told Madison to put the if-else statement after Ted asked the question. Then, Ophelia guided the girls through training a speech classification system with two categories: happy and sad. For each category, they listed synonyms for happy and sad that the user might use when replying to Ted. After creating the model, Ophelia left to help another group, and the girls attempted to integrate their conditional if-else statement with their speech classification model.

In this segment, Ophelia, Madison, and Amber engaged in programming practices and co-constructed knowledge about classification algorithms and the differences between AI applications that interact with real-world inputs and programming scripts that follow an algorithm regardless of outside stimuli. When creating their speech classification system, the girls added synonyms for happy and sad, thinking about two diverse users of Ted and how algorithms are made by humans and thus, their opinions and preferences of choosing "happy" and "sad" are embedded in their code.

4.2.6. Debugging a conditional statement

Madison and Amber succeeded at adding one condition to their algorithm which resulted in Ted responding to a user by playing happy music. However, Madison and Amber experienced challenges when adding a second condition to their if-else statement. Madison stated to Amber, "Yeah but I don't know what to put like right here". Madison was referring to the empty space under the "else" statement (Fig. 10). She tried to repeat what they had done with the if condition above and place the triangle shaped condition next to the "else". However, a condition did not need to be specified for an else statement; the girls simply needed to specify actions. Amber responded, "Yeah, how do we control it?" referring to her uncertainty around specifying a condition for their desired actions. Here, Amber used her own word of "control" to refer to debugging the code to reach their desired output. She pointed in the empty space, "Wait can you put it right there?" Madison unsuccessfully tried to fit the triangular block where Amber pointed. Madison externalized the problem, "We can't put, like, what we want for sad in here because we don't have, like, the other one, it's like a triangle". Amber noted, "It won't let us put anything in there". Amber and Madison used their everyday, anthropomorphizing language such as "it won't let us" and "it's like a triangle" to engage in debugging and problem solving in order to reach their desired goals.

They tinkered for another minute before Ophelia walked by again and asked, "How did it go guys?" Although facing a challenge, Madison optimistically stated, "Good! We are working on the sad part now". Ophelia looked at their code. She saw the if-else statement that they chose and, building on their existing code, explained that they could continue to use that block to achieve what they wanted. She explained that because they only have two conditions, by default, the else condition would be a response for sad. "You don't need an if-sad because the other option is sad right? So, if it's happy then do this, or else, do this other thing, which will be sad. So put all your sad stuff in there". Ophelia quickly walked away to assist another group.

4.2.7. Putting it all Together

Amber stated, "Okay. Let's do our stuff for sad", suggesting that they develop a series of actions that Ted would take when a user said they were sad. Madison suggested, "Play a sad song?" Amber replied, "That makes a lot more sense". Amber suggested that Ted should have blue headlights because "that would be a sad color". Madison agreed and added, "Okay drive forward. What else?" Amber suggested that in his sad condition, Ted should not spin around because that "kind of looks like its happy". The girls cycled through several rounds of tinkering and testing with the sad condition. Then, they tested both conditions. Madison exclaimed, "And I think we're done!" Amber agreed and added, "Yeah that's a pretty good program".

4.2.8. Summary: Engaging in computational practices when developing AI for social good

In this example of Madison and Amber working together with guidance from Ophelia to program Ted, there was evidence for all four sets of Scratch computational practices. Throughout the session, Madison and Amber *experimented and iterated* with their program. They tinkered with the blocks, tried out different code combinations, tested their program with Ted, and then revised their code to better meet their goals. When their program did not work as anticipated, they *tested and debugged* by identifying the problem and applying different strategies to solve the problem. For example, when Madison and Amber were experimenting with conditional statements, they identified a problem in which they

Fig. 10. Madison and Amber's code Scratch AI progression with a conditional if-else statement.

could not add a second condition. They programmed a "happy" condition but could not fit a desired condition into their if-else statement. As Amber put it, "It won't let us put anything in there". To solve this problem, they tried repeating their actions when they specified their condition in the if section of their statement. When that failed, another strategy was to ask for assistance from someone else. The girls invited Ophelia, a researcher, to help them debug their code. Ophelia reviewed their code, then suggested a solution. The girls tried that solution and were satisfied with the result. Madison and Amber also *reused and remixed* when they reworked their original design of Ted and translated their idea into a viable robot prototype. They also built on Ophelia's existing ideas when she suggested incorporating a speech classification system into their program. Ophelia also built on the girls' ideas when she suggested that they keep their if-else statements and assume that their second condition was a default "else" condition, rather than suggesting they change their block to two separate if-statements. Together, Ophelia and the girls negotiated and built on each other's ideas to program Ted to respond empathetically to the user. Finally, Madison and Amber *abstracted and modularized* when they separated the tasks of creating algorithms for Ted's actions and creating AI algorithms for Ted's ability to listen and speak. Then, they integrated these two tasks through a conditional statement enabling separate groups or modules of code.

5. Discussion

When engaging in CML activities rooted in constructionism design principles, children used some form of tool or object-to-think-with Holbert and Wilensky (2019), Kafai and Resnick (1996), Papert and Harel (1991) to (re)construct their understanding of machine learning through a critical lens. Some tools, such as Google Quick Draw, provided an opportunity to use a ML application and answer critical questions about its limitations. Other tools, such as the Google Teachable Machine, provided a structured approach to train a ML application and explore the mechanisms behind biased datasets. Scratch AI blocks and the micro:bit robot provided an exploratory space for remixing fantastical robot stories into prototypes for marginalized populations. Children also connected their ML content knowledge to the history of racism and sexism when they watched and discussed the Coded Bias trailer. By the end of the program, children made personal connections with their robot constructions, writing their own creative stories, and designing robots for children in need or for other marginalized populations. These personal connections contributed to a rich understanding of underlying CML concepts and guided learners towards a meaningful relationship with both the robot and the concepts (Wilensky, 1991).

Regarding the progression of learning, the findings suggest that as children progressed through the program, they made more connections with socio-political orientations and ML content. In the pre activities, children made no connections across the CML codes. In the three middle activities, children made increasingly more sophisticated connections. When developing a prototype of their robot for social good, a detailed Discourse analysis suggested that children engaged in the four sets of computational practices, as outlined by Brennan and Resnick (Brennan & Resnick, 2012). Finally, in their post activities, children made the most connections across all the CML knowledge elements, suggesting a more integrated understanding of socio-political orientations and machine learning content by the end of the program. Specifically, children answered critical questions related to AI, such as Who develops technologies? For whom are technologies developed? and What decisions are made based on the outputs of the algorithms? Children's responses contained discussions of how dominant populations create the majority of technologies and that women and people of color may be unjustly excluded or harmed when biased datasets are used to train ML applications. In their robot stories, some children designed robots that could do broader social good such as build homes for those in need and pick up trash for recycling. Others focused on designing for specific marginalized populations such as other children who are not being served by current technologies. For example, LaToya designed a color recognition application for children, such as her cousin, who are displaced frequently and may not have access to learning tools of that kind. She used her trained Google Teachable Machine to create a prototype of this learning tool. In another example, although he ultimately did not create a prototype, Lucas imagined creating a set of headphones for autistic children to address sensory issues in public spaces. In line with the goals of this special issue, these findings demonstrate that children unpacked inequitable sociopolitical aspects behind algorithm bias. Thus, they increased their critical consciousness of issues that affect them and other excluded populations. With these opportunities to discuss inequities in algorithms, children uncovered historical injustices, and in their designer role, they proposed more equitable ways to correct them.

Thus, the set of findings in this study advance characterizations of justice-centered learning design by showing children's development of "pieces" of a critical lens by engaging in different aspects of critical thinking throughout the program. At times, children questioned their and others' roles in asymmetrical power structures and challenging unfair social structures that shape people's lives (Freire, 1970; Giroux, 1985). At other times, children saw a "view from somewhere" (Haraway, 1988), reflecting on the human designers of AI, discussing the oppressive consequences of biased training datasets in ML, and engaging in the socio-political contexts of technologies (Blikstein & Blikstein, 2021; Vakil, 2014; Vakil & Ayers, 2019) by designing innovative ML technologies for those whose perspectives have been historically neglected. Our justice oriented CML approach is in contrast to other educational approaches in computer science that take a narrow, individual view on ethics and to those that render the human designer invisible. In our approach, youth not only created products using already existing tools, but they also took an action-driven stance to imagine and design more equitable alternatives resisting dominant practices in algorithm bias.

Thus, the two main claims in this paper are that a critical lens to ML education can be characterized by 1) posing and answering questions about the roles of producers and consumers of AI technologies, specifically who designs technologies, for what purposes, who benefits, who is harmed, and what are the histories embedded in the data being used, and 2) identifying how people design AI technologies and applying this knowledge to build applications for marginalized populations. Taken together, these claims provide the foundations for characterizing children's development of conceptual knowledge and practices in environments where a critical lens is applied to a constructionist design of a computer science education program. To continue our contribution to create more equitable spaces for youth design, additional cycles of design and analysis in multiple contexts are needed to provide more evidence to further test and refine the CML approach and integrations of critical pedagogies, constructionism, and computer science education.

One challenge that we encountered to fully support children's critical lens throughout the CML program was that the digital technologies themselves did not contain a critical component. In the first stages of the program, the researchers guided learners through the critical elements. However, during the robot design process, some of this guidance was removed, creating inconsistencies for children to continue reflecting critically on their designs. Upon examination of this outcome, we realized that the computational tools themselves needed to explicitly embed opportunities to engage children in socio-political reflection during their design. This way, children's critical consciousness will emerge and develop early in the program through interactions with adults and pre-designed elements embedded in the technologies. To this end, future contributions of our work will focus on ways in which critical elements are distributed among tools. This effort will assist children's increasing awareness of their sociopolitical reality during all stages of designing for marginalized populations, and thus, as Lee and Soep put it, "building towards something better" (Lee & Soep, 2016).

In line with this solution to our challenge, we will implement the following changes in future design iterations to advance this design-based study. To understand more deeply how children integrate a critical perspective while learning computer science content and practices, the design of the program must more fully integrate knowledge, practices, and critical orientations into each activity and tool. Children should begin by answering critical questions and ideally, by the end of the program, pose their own critical questions to answer and share their design with the public, with the goal of impacting perceptions and policies in the world. In addition, in this first iteration there was not enough emphasis on the histories of oppression that drive current oppression embedded in harmful AI technologies. Although children explained how people develop technologies and the consequences of the large-scale deployment of technologies, the CML educational program did not draw enough explicit connections to the history of oppression that women or people of color have faced and how those histories are linked to the design of harmful AI technologies, such as the facial recognition software discussed in the Coded Bias film trailer. In the future, such nuanced histories of marginalized populations must be connected explicitly to technology development and deployment for children to develop deeper critical perspectives. Finally, our data collection procedures must be improved such that we capture conceptual knowledge and practices at the same time through children's use of our digital tools. This data capture is critical for developing network models of children's sense-making that include connections between knowledge and practices to better understand the integration of a critical lens to the domain of computer science education.

6. Conclusion

The findings in this study characterize how children applied a critical lens when learning machine learning conceptual knowledge and computational practices. These characterizations support the approach of CML education, which argues that machine learning education cannot be separated from the social, historical, and political contexts in which people consume and produce technology. This study suggests that a critical lens is an effective approach towards engaging young children in designing their own machine learning applications in socially responsible ways. The exploration presented in this study is just one example of the multiple possible approaches towards engaging youth early on in their education about machine learning concepts and how to think critically about the social, ethical, and political issues around modern large scale algorithm deployment. Such educational research explorations are crucial for breaking the harmful tradition of technology development and consumption without a critical lens.

Selection and participation

All children in this study were enrolled in the community after-school program. The study took place at the after-school center in a community room. Data related to the study were collected after approval from the Institutional Review Board at Clemson University, following all the regulations and recommendations for research with children. Researchers obtained written consent from the parents/guardians of all child participants permitting the data collection. Children were informed about the data collection process and their participation in the study was completely voluntary. In addition, children were able to withdraw their consent for the data collection at any time without affecting their participation in the activity.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

We would like to thank all the wonderful children who allowed us into their after-school space and allowed us to learn from them. This work is funded in part by the National Science Foundation (ECR-2024965) and the Office of the Associate Dean for Research in the College of Education at Clemson University. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

References

Arastoopour Irgens, G. (2021). Connected design rationale: A model for measuring design learning using epistemic network analysis. *Instructional Science*, *49*(4), 561–587. http://dx.doi.org/10.1007/s11251-021-09551-8.

Arastoopour Irgens, G., et al. (2022). Designing with and for Youth: A Participatory Design Research Approach for Critical Machine Learning Education. *Educational Technology & Society*, Submitted for publication.

Bacchini, F., & Lorusso, L. (2019). Race, again: how face recognition technology reinforces racial discrimination. *Journal of Information, Communication and Ethics in Society*, *17*, 321–335. http://dx.doi.org/10.1108/JICES-05-2018-0050.

Blikstein, P. (2008). Travels in Troy with Freire: technology as an agent for emancipation. In C. Torres, & P. Noguera (Eds.), *Social justice education for teachers: Paulo Freire and the possible dream, sense, Rotterdam, Netherlabds* (pp. 205–244). http://dx.doi.org/10.1163/9789460911446.

Blikstein, P., & Blikstein, I. (2021). Do educational technologies have politics? In M. Ito, & et al. (Eds.), *A semiotic analysis of the discourse of educational technologies and artificial intelligence in education*. MIT Press, http://dx.doi.org/10.1162/ba67f642.646d0673.

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

Bowman, D., Swiecki, Z., Cai, Z., Wang, Y., Eagan, B., Linderoth, J., et al. (2021). The mathematical foundations of epistemic network analysis. In *International conference on quantitative ethnography* (pp. 91–105). Springer.

Braden, E. (2019). Do you have papers?: Latinx third graders analyze immigration policy through critical multicultural literature. *Children's Literature in Education*, *50*, 464–480. http://dx.doi.org/10.1007/s10583-018-9359-5.

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 annual meeting of the American educational research association*. Canada: Vancouver, (1, p. 25).

Bucher, T. (2017). The algorithmic imaginary: exploring the ordinary affects of facebook algorithms. *Information, Communication & Society*, *20*, 30–44. http://dx.doi.org/10.1080/1369118X.2016.1154086.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77–91). http://proceedings.mlr.press/v81/buolamwini18a.html (accessed July 26, 2021).

Charmaz, K. (2008). Constructionism and the grounded theory method. In J. Holstein, & J. Gubrium (Eds.), *Handbook of constructionist research* (pp. 397–412). New York: The Guilford Press.

Dasgupta, S., & Hill, B. (2021). Designing for critical algorithmic literacies. In *Algorithmic rights and protections for children*. https://wip.mitpress.mit.edu/pub/designing-for-critical-algorithmic-literacies/release/1. (Accessed 6 September 2022).

D'Ignazio, C., & Klein, L. (2020). *Data feminism*. Cambridge, Massachusetts: The MIT Press.

Druin, A. (2002). The role of children in the design of new technology. *Behaviour and Information Technology*, *21*, 38.

Duberry, J. (2022). *AI in public and private forms of surveillance: challenging trust in the citizen-government relations*. Edward Elgar Publishing, https://www.elgaronline.com/view/book/9781788977319/book-part-9781788977319-9.xml. (Accessed 6 September 2022).

Dunham, Y., Baron, A., & Carey, S. (2011). Consequences of 'minimal' group affiliations in children. *Child Development*, *82*, 793–811. http://dx.doi.org/10.1111/j.1467-8624.2011.01577.x.

Fain, J. (2008). Um, they weren't thinking about their thinking: Children's talk about issues of oppression. *Multicultural Perspectives*, *10*, 201–208. http://dx.doi.org/10.1080/15210960802526102.

Fox-Skelly, J., Bird, E., Jenner, N., Winfield, A., Weitkamp, E., & Larbey, R. (2020). *The ethics of artificial intelligence: issues and initiatives*. LU: Publications Office of the European Union, https://data.europa.eu/doi/10.2861/6644 (accessed April 22, 2022).

Frary, M. (2017). Your cover is shown: Tech giants and governments are out to get your data. In *Soon it might be impossible to remain anonymous. Vol. 46* (pp. 39–41). Index on Censorship.

Freire, P. (1970). *Pedagogy of the oppressed, 30th anniversary ed, continuum*. New York.

Gee, J. (2008). *Social linguistics and literacies: ideology in discourses* (3rd ed.). London, New York: Routledge.

Gee, J. (2011). *An introduction to discourse analysis: theory and method* (3rd ed.). Milton Park, Abingdon, New York: Routledge.

Giroux, H. (1985). Critical pedagogy, cultural politics and the discourse of experience. *Journal of Education*, *167*, 22–41. http://dx.doi.org/10.1177/002205748516700204.

Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. *Educational Researcher*, *42*, 38–43.

Guha, M., Druin, A., & Fails, J. (2013). Cooperative inquiry revisited: Reflections of the past and guidelines for the future of intergenerational co-design. *International Journal of Child-Computer Interaction*, *1*, 14–23. http://dx.doi.org/10.1016/j.ijcci.2012.08.003.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, *14*, 575–599. http://dx.doi.org/10.2307/3178066.

Hautea, S., Dasgupta, S., & Hill, B. (2017). Youth perspectives on critical data literacies. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 919–930). Denver Colorado USA: ACM, http://dx.doi.org/10.1145/3025453.3025823.

Holbert, N., & Wilensky, U. (2019). Designing educational video games to be objects-to-think-with. *Journal of the Learning Sciences*, *28*, 32–72. http://dx.doi.org/10.1080/10508406.2018.1487302.

Kafai, Y. (2016). From computational thinking to computational participation in K–12 education. *Communications of the ACM*, *59*, 26–27.

Kafai, Y., & Resnick, M. (Eds.), (1996). *Constructionism in practice: designing, thinking, and learning in a digital world* (1st ed.). Mahwah, N.J: Routledge.

Kantayya, S. (2020). *Coded Bias [Film]. 7th Empire Media*. https://www.codedbias.com/.

Kersten, J. (2006). Why's everyone white? Moving toward critical pedagogy in an elementary classroom. *Journal of Urban Learning, Teaching, and Research*, *2*, 35–44.

Kramer, A., Guillory, J., & Hancock, J. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*, 8788–8790. http://dx.doi.org/10.1073/pnas.1320040111.

Lee, C., Gobir, N., Gurn, A., & Soep, E. (2022). In the black mirror: Youth investigations into artificial intelligence: In the black MirrorYouth investigations into artificial intelligence. *ACM Transactions on Computing Education*, Article 3484495. http://dx.doi.org/10.1145/3484495.

Lee, C., & Soep, E. (2016). None but ourselves can free our minds: Critical computational literacy as a pedagogy of resistance. *Equity & Excellence in Education*, *49*, 480–492. http://dx.doi.org/10.1080/10665684.2016.1227157.

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–16). Honolulu, HI, USA: ACM, http://dx.doi.org/10.1145/3313831.3376727.

Marquart, C., Hinojosa, C., Swiecki, Z., & Shaffer, D. (2018). ENA webtool. https://app.epistemicnetwork.org/login.html. (Accessed 30 July 2021).

O'Neil, C. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy* (1st ed.). New York: Crown.

Papert, S. (1980). Mindstorms: children, computers, and powerful ideas. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=982028. (Accessed 18 April 2022).

Papert, S., & Harel, I. (1991). *Constructionism*. Ablex Publishing.

Pea, R. (1993). Practices of distributed intelligence and designs for education. In G. Salomaon (Ed.), *Distributed cognitions: psychological and educational considerations*. New York: Cambridge University Press.

Pirrung, M., Hilliard, N., Yankov, A., O'Brien, N., Weidert, P., Corley, C., et al. (2018). Sharkzor: Interactive deep learning for image triage, sort and summary. [Cs] arXiv:1802.05316. http://arxiv.org/abs/1802.05316. (Accessed 22 April 2022).

Raub, M. (2018). Bots, bias and big data: Artificial intelligence, algorithmic bias and disparate impact liability in hiring practices. *Arkansas Law Review*, *71*(2), 529–570, https://scholarworks.uark.edu/alr/vol71/iss2/7.

Rosebery, A., Ogonowski, M., DiSchino, M., & Warren, B. (2010). The coat traps all your body heat: Heterogeneity as fundamental to learning. *The Journal of the Learning Sciences*, *19*, 322–357.

Samoili, S., Cobo, M., Delipetrev, B., Martinez-Plumed, F., Gomez, E., & Prato, G. (2021). AI watch. Defining artificial intelligence 2.0. Towards an operational definition and taxonomy of AI for the AI landscape. In *Joint research centre*. https://ideas.repec.org/p/ipt/iptwpa/jrc126426.html. (Accessed 18 April 2022).

Shaffer, D. (2017). *Quantitative ethnography*. Madison, WI: Cathcart Press.

Shaffer, D. (2018). Epistemic network analysis: Understanding learning by using big data for thick description. In F. Fischer, C. Hmelo-Silver, S. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 520–531). New York: Routledge, http://dx.doi.org/10.4324/9781315617572.

Shaffer, D., & Ruis, A. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In *Handbook of learning analytics*.

Shamir, G., & Levin, I. (2022). Teaching machine learning in elementary school. *International Journal of Child-Computer Interaction*, *31*, Article 100415. http://dx.doi.org/10.1016/j.ijcci.2021.100415.

Smith, M., & Miller, S. (2022). The ethical application of biometric facial recognition technology. *AI & Society*, *37*, 167–175. http://dx.doi.org/10.1007/s00146-021-01199-9.

Stark, L. (2019). Facial recognition is the plutonium of AI. *XRDS*, *25*, 50–55. http://dx.doi.org/10.1145/3313129.

Stornaiuolo, A. (2020). Authoring data stories in a media makerspace: Adolescents developing critical data literacies. *Journal of the Learning Sciences*, *29*, 81–103. http://dx.doi.org/10.1080/10508406.2019.1689365.

The European Commission (2019). A definition of AI: Main capabilities and scientific disciplines, high-level expert group on artificial intelligence. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence.

Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI? *AAAI*, *33*, 9795–9799. http://dx.doi.org/10.1609/aaai.v33i01.33019795.

Vakil, S. (2014). A critical pedagogy approach for engaging urban youth in mobile app development in an after-school program. *Equity & Excellence in Education*, *47*, 31–45. http://dx.doi.org/10.1080/10665684.2014.866869.

Vakil, S., & Ayers, R. (2019). The racial politics of STEM education in the USA: interrogations and explorations. *Race Ethnicity and Education*, *22*, 449–458. http://dx.doi.org/10.1080/13613324.2019.1592831.

Wilensky, U. (1991). Abstract meditations on the concrete and concrete implications for mathematics education. In *Constructionism* (pp. 193–203). Westport, CT, US: Ablex Publishing.

Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: learning biology through constructing and testing computational theories—An embodied modeling approach. *Cognition and Instruction*, *24*, 171–209. http://dx.doi.org/10.1207/s1532690xci2402_1.

Williams, R. (2021). How to train your robot: Project-based AI education for middle school classrooms. In *SIGCSE '21: ACM special interest group on computer science education*. Toronto: Association for Computing Machinery.