"How Can I Code A.I. Responsibly?": The Effect of Computational Action on K-12 Students Learning and Creating Socially Responsible A.I.

H. Nicole Pang, Robert Parks, Cynthia Breazeal*, Hal Abelson*

Massachusetts Institute of Technology
77 Massachusetts Avenue Cambridge, MA 02139
hpang@alum.mit.edu, rparks@mit.edu, cynthiab@media.mit.edu, hal@mit.edu

Abstract

Teaching young people about artificial intelligence (A.I.) is recognized globally as an important educational effort by organizations and programs such as UNICEF, OECD, Elements of A.I., and AI4K12. A common theme among K-12 A.I. education programs is teaching how A.I. can impact society in both positive and negative ways. We present an effective tool that teaches young people about the societal impact of A.I. that goes one step further: empowering K-12 students to use tools and frameworks to create socially responsible A.I. The *computational action process* is a curriculum and toolkit that gives students the lessons and tools to evaluate positive and negative impacts of A.I. and consider how they can create beneficial solutions that involve A.I. and computing technology. In a human-subject research study, 101 U.S. and international students between ages 9 and 18 participated in a one-day workshop to learn and practice the computational action process. Pre-post questionnaires measured on the Likert scale students' perception of A.I. in society and students' desire to use A.I. in their projects. Analysis of the results shows that students who identified as female agreed more strongly with having a concern about the impacts of A.I. than those who identified as male. Students also wrote open-ended responses to questions about what socially responsible technology means to them pre- and post-study. Analysis shows that post-intervention, students were more aware of ethical considerations and what tools they can use to code A.I. responsibly. In addition, students engaged actively with tools in the computational action toolkit, specifically the novel *impact matrix*, to describe the positive and negative impacts of A.I. technologies like facial recognition. Students demonstrated breadth and depth of discussion of various A.I. technologies' far-reaching positive and negative impacts. These promising results indicate that the computational action process can be a helpful addition to A.I. education programs in furnishing tools for students to analyze the effects of A.I. on society and plan how they can create and use socially responsible A.I.

Introduction

Teaching young people about artificial intelligence (A.I.) is recognized globally as an important education effort by organizations and programs such as UNICEF, OECD, Elements of A.I., and AI4K12. Over the last few years, these and other organizations have also prioritized classroom discussions of the ethics of A.I., reflecting the field's growing impact on students' everyday lives. Anecdotally, teachers report that classroom discussions of ethics are a highly engaging means of building critical thinking skills by weighing benefits against harms and reflecting on impacts on various stakeholders. "Incorporating ethical or social discussions into science and engineering courses can increase student retention, particularly underrepresented minorities and women" (Lee et al. 2021). However, researchers and educators have noted that ethical components in current K-12 curricula are often sparse (Zhou, Van Brummelen, and Lin 2020) or treat students only as users of the technologies (Vakil 2018) (with topics such as "Keeping your data private" or "Looking out for biased recommendations in your social media feed"). A design-centric view of A.I. ethics asks students to consider design choices to mitigate systemic societal issues such as job loss or discrimination and employ empathy to envision A.I.'s impact on various stakeholders.

We present work that provides substantive opportunities for student engagement in responsible A.I. by scaffolding written and oral reflection as part of a novel computational action curriculum. The curriculum, called the *computational action process*, is based on the framework of Tissenbaum, Sheldon, and Abelson (Tissenbaum, Sheldon, and Abelson 2019). We conducted a human-subject research study with 101 U.S. and international students between ages 9 and 18 who participated in a one-day workshop to learn and practice the computational action process. Pre- and post-questionnaires deployed during the research measured on the Likert scale students' perception of responsible A.I. in society and students' desire to use A.I. responsibly in their projects. Analysis of the pre-study data shows that students who identified as female had a statistically significant higher concern about the impacts A.I. than those who identified as male. Analysis also showed some differences in concern about the impacts A.I. and interest in using A.I. between students from the U.S. and students from international countries. Students also wrote open-ended responses to questions about socially responsible technology's meaning to them pre- and post-study. Analysis of open-ended answers shows that more students discussed the ethics of technology post-intervention. In addition, students engaged with tools in the computational action toolkit, specifically the *impact*

*These authors contributed equally.

matrix, to discuss and reflect on the positive and negative impacts of A.I. technologies such as facial recognition and eye-tracking. Students demonstrated breadth and depth of discussion of various A.I. technologies' far-reaching positive and negative effects. These promising results indicate that the computational action process can be a helpful addition to A.I. education programs in furnishing tools that allow students to critically engage in issues of concern as they plan how to create and use socially responsible A.I.

Background and Related Work

Many recent curricula for A.I. promote a Project-Based Learning (PBL) methodology (Ng et al. 2021) as a means for students "to understand the working principles of A.I.-based systems by developing them, not only by using them" (Bellas et al. 2022). A common way to structure formal and informal PBL in computer science is through the engineering design process (Ng et al. 2021), scaffolding students through project phases from scoping a problem to iterating on a prototype as a possible solution (Moore et al. 2014).

We situate the computational action process within these hands-on methodologies with some pedagogical distinctions. Many PBL curricula teaching A.I. make use of a "design challenge" format or specified activity, in which problem-definition is supplied to students or student-teams during their design process. In contrast, we posit that an open-ended design process in which students name their own goals provides enhanced opportunities for the evaluation of ethical consequences of A.I. The computational action process is intended to scaffold the often-difficult task of generating unique project ideas and evaluating their implications. When successful, students given the agency to lead their own problem definition can feel more accountable for the consequences of their designs and more interested in viewing them as live questions through an ethical lens (Druga et al. 2019).

Various previous work also provides evidence for greater persistence and deeper engagement using a student-led engineering design process. Papert's constructionism (Harel and Papert 1991) and related work documents higher levels of task persistence and interest in learning when undertaking activities that are "personally meaningful" to students (Lodi and Martini 2021). "Personally meaningful" can be unpacked in recent work such as "goal congruency theory" (Diekman and Steinberg 2013), which contends that people are motivated to pursue careers and activities consistent with their social roles. Such roles defined culturally, personally, or historically can often clash with the perceived goals of a field of study such as engineering. For example, a student who thinks of themselves as socially minded may need to see alignment with those goals in existing technologies or reflect that they are free to forge such goals in their own design process within the field.

Research on student reflection, whether in class discussions or written responses, factors into this work. Reflection on the real harms and benefits of technologies, products, and ideas in the STEM education fields has resulted in improved student retention and engagement in these fields (Oskotsky et al. 2022; Yeager et al. 2014). Students are more likely to connect with engineering tasks impacting their communities or themselves (Castaneda and Mejia 2018; Ryoo 2019) and to debate issues of concern as they take on design and development roles for their ideas. Dynamic, critical reflection, beyond enumerating benefits in an "A.I. for Good"-style program, makes A.I. technologies less abstract and connects students with their own goals.

Several other curricula and frameworks have advanced the use of student reflection on ethics in A.I. (Holmes et al. 2022; Touretzky et al. 2019; Long and Magerko 2020; Hsu, Abelson, and Van Brummelen 2022). Touretzky's "Five Big Ideas of A.I.," promulgated by the organization AI4K12, establishes "Societal Impact" as the fifth, overarching principle: "Students should understand that the ethical construction of A.I. systems that make decisions affecting people's lives requires attention to the issues of transparency and fairness" (Touretzky et al. 2019). As a means of preparation for future careers, Long and Magerko have identified ethics as a computer science competency in preparing future coders to consider how their designs will be used in the context of privacy, fair employment, transparency, and other factors (Long and Magerko 2020). Critical thinking exercises on responsible A.I. can be useful for students as citizens and potential professionals in the field and are often one of the most engaging parts of a curriculum. A recent A.I. literacy intervention (Lee et al. 2021) with 31 middle school-aged children of diverse backgrounds found that ethics-related activities (named "Unanticipated Consequences" and "Investigating Bias in A.I.") tested highest among materials according to coded student interviews.

Computational Action Process

The computational action process presented in this paper was created to put an ethical and impact-driven lens for students creating technology (i.e., programming apps or coding A.I. technology) and elaborates on the components of computational identity and digital empowerment laid out in the computational action framework (Tissenbaum, Sheldon, and Abelson 2019). The process was also influenced by industry-standard engineering design practices and tailored to meet K-12 standard expectations. The computational action process teaches students five topics:

- Defining a real-world problem
- Understanding users and communities
- Designing responsibly with users and communities
- Teamwork, project management, and implementation
- Making a long-lasting impact

Students practice each topic of the process with the tools in the computational action toolkit, which consists of:

- Mind map for brainstorming meaningful problems
- User research template, user persona template, and collaborative analysis framework
- Impact matrix, feature importance vs. cost tool, and tools for wireframing design
- Teamwork task management table, project management board
- Project reflection matrix, future timeline plan

The full set for each topic is open-source, and available at: https://docs.google.com/presentation/d/1AiD-r81_abJkJG_mLidS2yribn5ZRH8InP4jOS5-tMc. The full computational action toolkit is also open-source, and available at: https://drive.google.com/drive/folders/1aXN1QMVaN72QwUCJOosbzYHnuXRCOGbf.

Responsible A.I. Lessons

The computational action process introduced students to an impact-driven, ethics-based engineering design practice via facilitation, discussion, interactive online tools, and framing to understand and evaluate socially responsible technology and responsible A.I. The first topic of computational action helped students discover a real-world problem they care about. The curriculum covering this topic introduced students to the United Nations Sustainable Development Goals (UN SDG) for inspiration from real-world issues, and tools they can use to observe the community and people around them to find a personally meaningful problem to work on. Topic two introduced students to why user research is important when making something with A.I. The lesson taught best practices and concrete steps to conduct user and community research. The curriculum materials were created by the researchers, who have extensive backgrounds working in product management in the industry at companies such Google; real-world engineering design experiences and other general design resources inspired the curriculum.

They Topic three was an essential element for investigating ethics of A.I. and how to design and create responsible A.I. The *impact matrix* was developed as a bespoke tool to help students reason about the impact of technology solutions (both positive and negative) and understand ethical A.I. Students observed example impact matrices for several contemporary technology issues involving A.I., such as social robots and screen monitoring for at-home schooling. They were introduced to methods for weighing the positives and negatives of A.I. technology with regard to their impact on society. Stakeholders' values were discussed, including typical values such as privacy, security, safety, and accessibility. The rest of the third topic covered how to create designs for their project proposals, which covered sketching, wireframing, fast prototyping, and testing with real users.

The last two topics of computational action focused on implementation, reflection, and the iterative process of continuing to get feedback from users. Topic four gave students real-world tools for implementation and project management. Topic five emphasized the cyclical nature of the computational action process by encouraging students to plan for future additions to their A.I. projects after they get real user feedback. These materials allow students to engage in a critical pedagogy to enhance connections to meaningful and authentic practice.

Students engaged with these lessons and materials through instruction taught by the lead researcher. The specific tools and activities of the intervention that taught and guided students to learn and discuss responsible A.I. appear in Fig. 1.

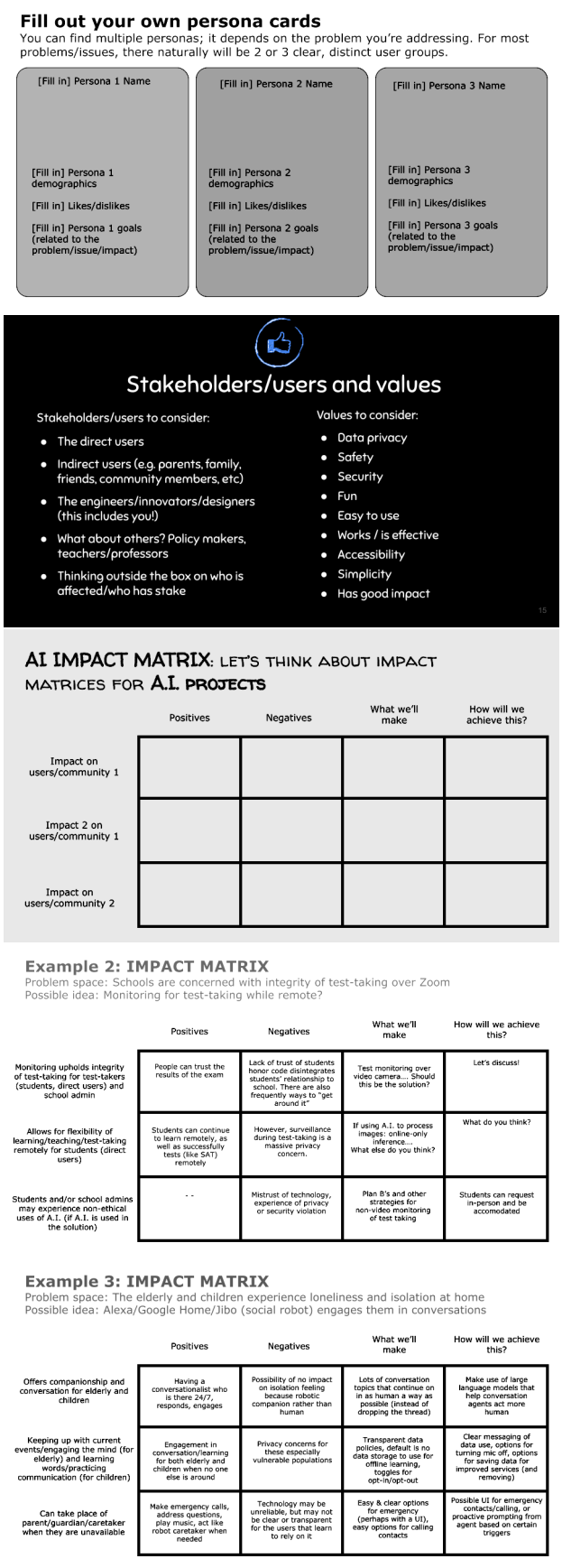

Figure 1: Some of the tools and activities regarding responsible A.I., as seen by students (original in color).

Question Type	Question Number	Question
Perception of responsible A.I.	Q14	I want to include artificial intelligence (A.I.) in technology projects I create
Perception of responsible A.I.	Q15	I am concerned about the use of artificial intelligence (A.I.) in technology

Figure 2: Two questions on responsible A.I. in the pre-post survey instrument.

Method

The research questions we investigated were: (1) What interventions change students' perceptions of A.I. technology in their community? and (2) Is the computational action process effective in changing students' interest in making a positive impact using A.I. technology? The intervention consisted of a one-day online workshop to teach students all five topics of the computational action process and practice the computational action toolkit.

101 participants between the ages of 9 to 18 were recruited from mailing lists of US and international students associated with Technovation Girls, MIT Education Studies Program (ESP), and Solv(Ed) programs. The study was designed with two cohorts: cohort 1 consisted of students who have been previously introduced to coding and elements of engineering design, and cohort 2 consisted of students who have not been in these types of programs. As much as possible, the other variables between the two cohorts were kept constant. The research study protocol was approved by the Institute Review Board (IRB) associated with the researchers' institution. Participants from both cohorts participated in the one-day workshop online over the Zoom video conferencing platform. Students were asked to complete pre-post survey questions on the Likert scale and an open-answer pre-post questionnaire, which served as the measurements for the study.

Participants Out of 101 total participants from both cohorts who completed the pre-study survey, 58 identified as female, 42 identified as male, and 1 participant identified as non-binary. 59 participants were located in the U.S., 10 were from Lebanon, and 7 were located in India. Of the 101 participants, 24 were of age 12, 37 were of age 13, 11 were of age 14, and 10 were of age 15.

Data Collection and Analysis Participants in both cohorts received the same pre-post survey questions, scored on the Likert scale from 1 (strongly disagree) to 5 (strongly agree). Participants were instructed to complete the pre-study questionnaire before the workshop intervention, and complete the post-study questionnaire immediately after. The survey questions used can be seen in Fig. 2.

The analysis of quantitative survey data was done using tests corresponding to the data distribution (whether normal or not normally distributed). Paired tests compared pre-post data of the same individuals, and unpaired tests compared different segments of either pre- or post-data (e.g., female vs. male responses). For paired results, data that followed

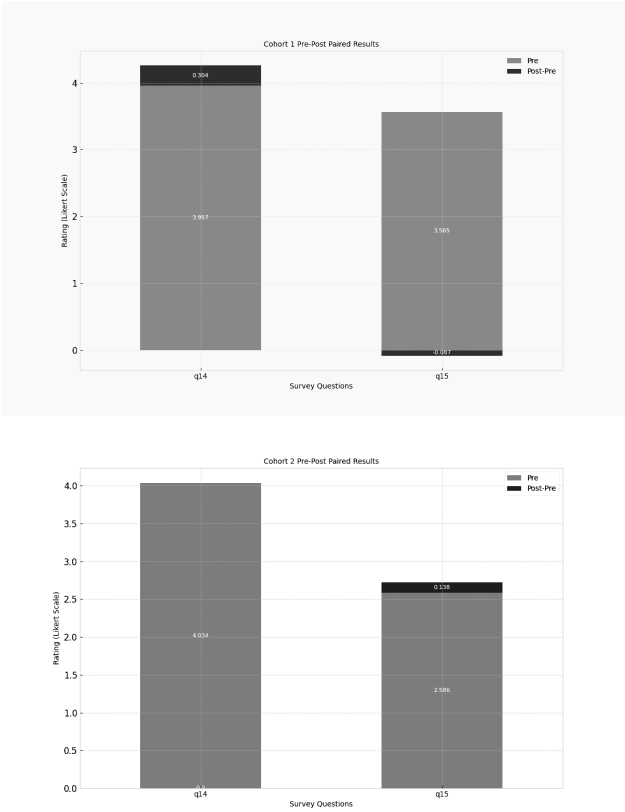

Figure 3: Top: plot of cohort 1 pre-post paired results; bottom: plot of cohort 2 pre-post paired results.

normal distribution were analyzed using paired t-test; otherwise, non-normally distributed data were analyzed using the Wilcoxon signed-rank test. For unpaired results, data that followed normal distribution were analyzed using a two-group t-test, and data that were not normally distributed were analyzed using the Mann-Whitney U-test. P-value of 0.05 determined whether results were significant.

Results

Analysis of paired pre-post survey responses from both cohorts of students showed no significant change pre-post intervention to both of the "Perception of Responsible A.I." questions in the survey. The paired results can be seen in Fig. 3. Analysis of unpaired results from the quantitative survey instrument showed some significant differences between certain independent variables. The unpaired results analyzed the results from both cohorts. The following differences were significant.

Pre-Intervention Unpaired Results From the pre-study survey, participants who identified as female agreed more strongly with having concerns about A.I. (Q15) than participants who identified as male (Female/Male: $\bar{x}$=3.172,2.667; p=0.046; t(100)=2.02). On perception of A.I., students from Lebanon more strongly agreed with having concerns about A.I. than students from the U.S. (US/Lebanon: $\bar{x}$=2.73,3.6;

p=0.039; U(69)=178.5). Analysis also showed that participants of age 15 had a stronger concern about the use of A.I. in technology than participants of age 12 (Q15 Age 12/Age 15:$\bar{x}$=2.375,3.7; p=0.0039; U(34)=45.5).

On their interest in using A.I. in their own projects (Q14: "I want to include artificial intelligence (AI) in technology projects that I create"), students from India ranked their interest more strongly than students from U.S. in the pre-survey (US/India: $\bar{x}$=3.847,4.714; p=0.0387; U(66)=111.5).

Post-Intervention Unpaired Results Of the 65 participants who completed the post-study survey from both cohorts, 21 participants were of age 13, 17 were of age 12, 6 participants were of age 14, 8 of age 15, and 4 of age 16. Participants of age 15 indicated a greater concern about the use of artificial intelligence (A.I.) in technology than participants of age 12. (Age 12/Age 15: $\bar{x}$=2.294,4; p=0.00237; U(25)=17). None of the other independent variables, when analyzed on post-study survey results for the questions on responsible A.I., demonstrated significant change.

A.I. Impact Matrix Results the mean In the study, the instructor introduced the *impact matrix* to students and then guided students in a group discussion and activity, which used the impact matrix to examine the meaning of ethical A.I. The problem posed for the group activity centered on using A.I. for quicker and better stroke detection by using facial recognition to detect asymmetry. Students from cohort 2 separated into two random groups to do group activities. Both groups of students proactively shared deep and insightful discussion of potential harms of A.I. technology, but also potential life-changing benefits. They discussed positive impacts and negative side effects or harms with each other extensively through sharing aloud and the text chatbox provided by the Zoom video conferencing platform. Only lightly guided by the instructor, students produced insightful, deep discussion and ideas for solutions that are mindful of negative consequences on users. The results of the A.I. impact matrices created jointly by the students are in Fig. 4.

Qualitative Results

Students in cohort 2 responded to questions on their perceptions of making socially responsible technology with short, written answers. We performed an inductive analysis of the responses to identify themes and codes related to socially responsible technology. Two researchers iteratively developed the codes, then convened to discuss the code results. The qualitative results from these questions provide another means of assessing the intervention's effect on understanding and motivation to create socially responsible technology.

Pre-Intervention Responses 79% of students in cohort 2 answered the pre-survey question ("What does socially responsible technology in society mean to you?"). The highest number of students (41%) responded with the theme of promoting non-specific ethical benefits: in other words, making the world better. For example, student P9 wrote: "To me that means technology that impacts out society in a positive way." P17 stated: "Not quite sure what you mean, but

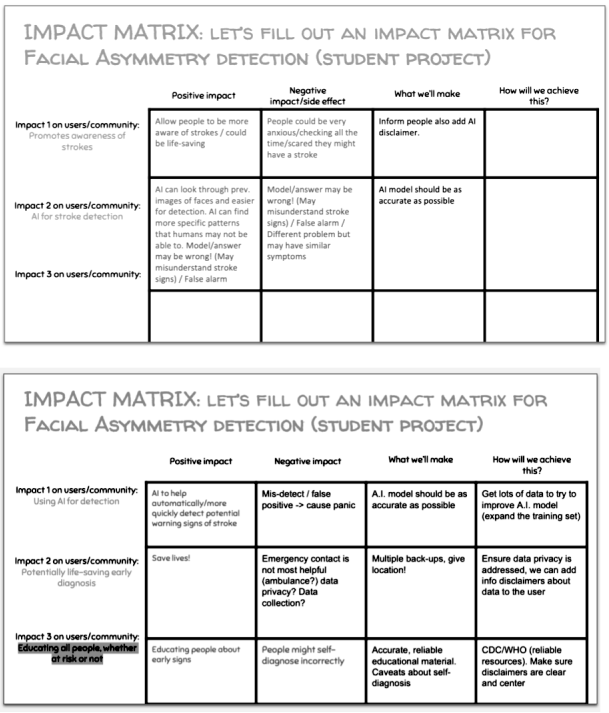

Figure 4: The two impact matrices that students from cohort 2 created jointly, discussing with insight and detail the benefits and harms of A.I.

Pre-workshop responses to "What does socially responsible technology in society mean to you?"

Themes:

- Promoting specific social benefits (10%)
- Promoting non-specific social benefits (41%)
- Preventing specific social harms (15%)
- Preventing non-specific social harms (7%)
- Using ethical considerations (10%)
- Don't know (17%)

Post-workshop responses to "After this class, what does socially responsible technology now mean to you?"

Themes:

- Promoting specific social benefits (0%)
- Promoting non-specific social benefits (34%)
- Preventing specific social harms (9%)
- Preventing non-specific social harms (0%)
- Using ethical considerations (31%)
- Don't know (25%)

Table 1: Themes resulting from the qualitative survey questions asked pre- and post-intervention (% of responses)

I guess that's the point, perhaps like responsible technology, like creating technology that's humane and has an obligation to be...socially responsible?" Some responses within

Theme: Promoting specific social benefits
• Socially responsible technology is private/trustworthy (5%)
• . . . helps the environment (5%)
Theme: Promoting non-specific social benefits
• . . . helps society (34%)
• . . . helps specific communities (7%)
Theme: Preventing specific social harms
• . . . does no harm via social media (15%)
Theme: Preventing non-specific social harms
• . . . does no harm to society (7%)
Theme: Using ethical considerations
• . . . is ethically designed (10%)
Theme: Don't know
• Don't know (17%)

Table 2: Pre-workshop response codes (% of responses)

Theme: Promoting specific social benefits
• Socially responsible technology is private/trustworthy (0%)
• . . . helps the environment (0%)
Theme: Promoting non-specific social benefits
• . . . helps society (19%)
• . . . helps specific communities (16%)
Theme: Preventing specific social harms
• . . . does no harm via social media (9%)
Theme: Preventing non-specific social harms
• . . . does no harm to society (0%)
Theme: Using ethical considerations
• . . . is ethically designed (31%)
Theme: Don't know
• Don't know (25%)

Table 3: Post-workshop response codes (% of responses)

this theme were vague or rephrased the question.

A separate category of respondents cited specific harms as a focus of socially responsible technology (15%). All harms mentioned were related to social media, such as P6: "It means being as responsible with what you say and do as you are in real life because everything done on the internet doesn't leave once posted or sent." Those who described specific social benefits (10%) noted environmental actions and good privacy.

A portion of students (10%) framed their response in terms of ethics or an ethical framework, such as P33: "It means that the use of technology benefits society more than it harms it" or P38: "Technology that improves people's lives without hurting others or the environment."

Post-Intervention Responses 64% of students in cohort 2 answered the post-survey question. Just as with the pre response, the largest group of responses in post (35%) also focused on the non-specific social benefits of socially responsible technology. For example, student P38 wrote: "Socially responsible technology now means technology that benefits society positively based on what users need, want, and expect." Notably, a greater number within this theme (16% in post compared to 7% in pre) defined public good in terms of a community or communities, perhaps reflecting the workshop's focus on considering impacts on specific users and communities. P19 stated: "Technology that can not only help a community, but LEAD it, too!"

The second largest theme among responses (31%) defined socially responsible technology in terms of an ethical framework or used the term ethics. Student P3, for instance, wrote: "Technology which's positives out-weight the negatives." P23: "It means that it is an app or device that respects a human's boundaries." P40: "It means technology that makes a net positive impact in the community, region, country, and/or world." No discussion of specific harms or specific benefits were expressed.

Discussion

Data and student work show that the intervention improved K-12 students' awareness of using tools and frameworks to evaluate ethical considerations of their computational designs, including apps using A.I. technologies. The results of coded qualitative data in pre- and post-questionnaires show changes in how students approached the question: "What does socially responsible A.I. technology mean to you?" Before the intervention, many students (25%) described specific issues associated with specific technologies (cyberbullying, privacy, or environmental remedies) and wrote the perspective of a technology user ("Don't hide under a fake guise and be rude and mean to people and don't overdo or overuse technology.") (It is possible that references to privacy and cyberbullying may be the result of prior knowledge from media literacy classes in many schools.)

Post-intervention responses show that a larger number of respondents (31% vs. 10% of pre- responses) answered in terms of ethical and impact-based, user-centric considerations that could account for the benefits and harms of any A.I. or technology solution. ("With good impacts to your community, there can always be bad. That is why I have to be careful about what I do to impact my community in a good way" and "I think about the process.") Student reflections became more process-oriented; students began to see ethical and social considerations beyond single consumer applications and view themselves as evaluating technologies as designers and creators. Students also discussed using resources

mentioned in the workshop as a way to evaluate ideas, such as the impact matrix and United Nations Sustainable Development Goals, and expressed more confidence in making assessments when designing technology solutions ("The class made me think that it is easier to do than previously thought.")

Other qualitative data, however, reflect the difficulty in facilitating students in thinking about designing responsible A.I. in such a short intervention. Analysis shows that over a quarter of students were less sure of their understanding of "socially responsible" A.I. technology after participation (25% responded "don't know" vs. 17% previously). It is possible that consideration of A.I. in the context of harms and benefits is too unfamiliar a topic and that the workshop did not adequately scaffold the topic in the available time.

Quantitative data show no significant change in answers pre- and post- to both questions ("I want to include artificial intelligence (A.I.) in technology projects I create." and "I am concerned about the use of artificial intelligence (A.I.) in technology.") Students agreed or strongly agreed that they were interested in using A.I. in their coding projects (great than 4 out of 5) both pre- and post-study, but also, on average, agreed with having a concern about A.I. in today's society both pre- and post-study. The two questions may likely not be enough to tease out changes in willingness to use A.I. and concerns in doing so. For example, did students now feel that their concerns regarding their own designs were still high after using an ethical impact matrix tool? Or, had their concerns in their own designs been mitigated by changes they made upon reflection?

Also interesting is that unpaired results from these data show that students who identified as female showed more concern about the use of AI than those who identified as male. It is possible this result echos data by Priniski and other researchers into misalignments between cultural or historically informed values of the student versus subjective perceptions of the field (Priniski et al. 2019; Wigfield and Eccles 2000). Differences in gender responses prompt areas of future study (and testing with more refined questions).

Future Work

Tools, group discussion, and facilitation provided by the computation action toolkit provide various means of reflection on ethical considerations for young designers of A.I. technology. Additional opportunities for ethical reflection would likely improve student ideas and enhance motivation and persistence in the activity itself. Written reflection is a particularly effective intervention (Yeager et al. 2014; Priniski et al. 2019) in helping students find the "Why" – the personal perception of the utility value of the task (Wigfield and Eccles 2000). Future versions of the process should insert more opportunities for reflecting on and researching potential benefits and harms of a design, which can be done over an intervention of longer duration, and adding more detailed stakeholder values in the context of A.I. technology.

Students may not have had enough background knowledge to enumerate relevant A.I. benefits and harms for their project. Although the workshop provided time for student discussion and exposed participants to examples of relatively advanced student work in A.I., it did not provide instruction on the specifics of A.I. benefits and harms. Future iterations of the curriculum can guide students to seek more information about harms (such as algorithmic bias, data collection bias, misinformation, GANs used to produce deepfakes) in an inquiry-based learning model. It is also worth investigating whether topics should be slightly modified depending on the needs of international students to inform ethical reflection from a global perspective.

In addition, future research can include more survey questions that further plumb the details of students' interest in coding A.I. responsibly and their concerns about A.I. The noticeable shifts in students' qualitative responses to their perception of responsible A.I. and socially responsible technology in their communities indicate that more detailed survey questions may reveal specific shifts in students' perceptions. A longer research study can also observe and examine real A.I. applications that students create using the computational action process toolkit. Investigation of students' coded A.I. artifacts can shed further light on the effectiveness of the computational action process in helping students shift into creators of responsible technology, including A.I. Finally, computational action process materials are open-source and available to all teachers. It would be worthwhile to study a longer intervention, such as a school semester, and whether this could further clarify practices in creating A.I. technologies.

Ethical statement. The computational action process and the research study presented here aim to teach responsible A.I. to K-12 students, both as informed citizens and as creators of A.I. technologies. We hope that this work and the open-source materials provide a lasting tool for students who are learning and coding A.I. to evaluate ethical implications on society and make choices toward helping society and mitigating harm.

Acknowledgments

We thank the MIT RAISE initiative, App Inventor Group, and Personal Robots Group for funding and support. Additionally, we thank collaborators who provided help during the research study including Selim Tezel, Karen Lang, Mike Tissenbaum, Josh Sheldon, Jessica Van Brummelen, and Sharifa Alghowinem. Thank you also to Technovation, MIT ESP, and SOLV(ED) programs for helping with recruiting participants for the study.

References

Bellas, F.; Guerreiro-Santalla, S.; Naya, M.; and Duro, R. J. 2022. AI Curriculum for European High Schools: An Embedded Intelligence Approach. *International Journal of Artificial Intelligence in Education*, 1–28.

Castaneda, D. I.; and Mejia, J. A. 2018. Culturally relevant pedagogy: An approach to foster critical consciousness in civil engineering. *Journal of Professional Issues in Engineering Education and Practice*, 144(2): 02518002.

Diekman, A. B.; and Steinberg, M. 2013. Navigating social roles in pursuit of important goals: A communal goal con-

gruity account of STEM pursuits. *Social and personality psychology compass*, 7(7): 487–501.

Druga, S.; Vu, S. T.; Likhith, E.; and Qiu, T. 2019. Inclusive AI literacy for kids around the world. In *Proceedings of FabLearn 2019*, 104–111.

Harel, I.; and Papert, S., eds. 1991. *Constructionism : research reports and essays, 1985-1990*. Cambridge, Mass.: Massachusetts Institute of Technology. Media Laboratory.

Holmes, W.; Porayska-Pomsta, K.; Holstein, K.; Sutherland, E.; Baker, T.; Shum, S. B.; Santos, O. C.; Rodrigo, M. T.; Cukurova, M.; Bittencourt, I. I.; et al. 2022. Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 32(3): 504–526.

Hsu, T.-C.; Abelson, H.; and Van Brummelen, J. 2022. The Effects on Secondary School Students of Applying Experiential Learning to the Conversational AI Learning Curriculum. *International Review of Research in Open and Distributed Learning*, 23(1): 82–103.

Lee, I.; Ali, S.; Zhang, H.; DiPaola, D.; and Breazeal, C. 2021. Developing Middle School Students' AI Literacy. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, SIGCSE '21, 191–197. New York, NY, USA: Association for Computing Machinery. ISBN 9781450380621.

Lodi, M.; and Martini, S. 2021. Computational thinking, between Papert and Wing. *Science & Education*, 30(4): 883–908.

Long, D.; and Magerko, B. 2020. "What is AI Literacy? Competencies and Design Considerations". In *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–16. New York, NY, USA: Association for Computing Machinery. ISBN 9781450367080.

Moore, T. J.; Glancy, A. W.; Tank, K. M.; Kersten, J. A.; Smith, K. A.; and Stohlmann, M. S. 2014. A framework for quality K-12 engineering education: Research and development. *Journal of pre-college engineering education research (J-PEER)*, 4(1): 2.

Ng, D. T. K.; Leung, J. K. L.; Chu, S. K. W.; and Qiao, M. S. 2021. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2: 100041.

Oskotsky, T.; Bajaj, R.; Burchard, J.; Cavazos, T.; Chen, I.; Connell, W. T.; Eaneff, S.; Grant, T.; Kanungo, I.; Lindquist, K.; et al. 2022. Nurturing diversity and inclusion in AI in Biomedicine through a virtual summer program for high school students. *PLoS computational biology*, 18(1): e1009719.

Priniski, S. J.; Rosenzweig, E. Q.; Canning, E. A.; Hecht, C. A.; Tibbetts, Y.; Hyde, J. S.; and Harackiewicz, J. M. 2019. The benefits of combining value for the self and others in utility-value interventions. *Journal of Educational Psychology*, 111(8): 1478.

Ryoo, J. J. 2019. Pedagogy that supports computer science for all. *ACM Transactions on Computing Education (TOCE)*, 19(4): 1–23.

Tissenbaum, M.; Sheldon, J.; and Abelson, H. 2019. From Computational Thinking to Computational Action. *Commun. ACM*, 62(3): 34–36.

Touretzky, D.; Gardner-McCune, C.; Martin, F.; and Seehorn, D. 2019. Envisioning AI for K-12: What should every child know about AI? In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 9795–9799.

Vakil, S. 2018. Ethics, identity, and political vision: Toward a justice-centered approach to equity in computer science education. *Harvard Educational Review*, 88(1): 26–52.

Wigfield, A.; and Eccles, J. S. 2000. Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, 25(1): 68–81.

Yeager, D. S.; Henderson, M. D.; Paunesku, D.; Walton, G. M.; D'Mello, S.; Spitzer, B. J.; and Duckworth, A. L. 2014. Boring but important: a self-transcendent purpose for learning fosters academic self-regulation. *Journal of personality and social psychology*, 107(4): 559.

Zhou, X.; Van Brummelen, J.; and Lin, P. 2020. Designing AI Learning Experiences for K-12: Emerging Works, Future Opportunities and a Design Framework.