

Introduction

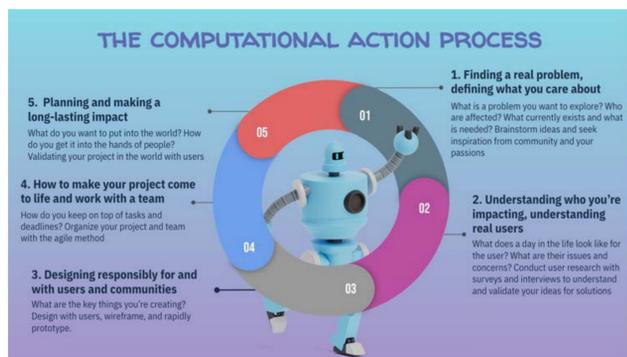
Students used the Computational Action toolkit to design socially responsible technology:

- Designing solutions based on an **impact matrix** to consider positives and negatives, as well as by sketching, wireframing, and testing
- Engaging in **critical reflection** scaffolded by the materials to make A.I. technologies less abstract and connecting students more deeply with their own goals

The Computational Action Process

The computational action process is:

- A curriculum and toolkit for middle and high school students
- Instead of “just coding,” students identify a meaningful problem they want to solve using A.I.
- Students aim to create socially responsible A.I. with lasting impact



Impact Matrix

The **impact matrix** was developed as a tool to help students reason about the impact of technology solutions (both positive and negative) and understand ethical A.I.

In the present study, students observed example impact matrices for several contemporary technology issues involving A.I., such as social robots and screen monitoring for at-home schooling.

Stakeholders/users and values

Stakeholders/users to consider:

- The direct users
- Indirect users (e.g. parents, family, friends, community members, etc)
- The engineers/innovators/designers (this includes you!)
- What about others? Policy makers, teachers/professors
- Thinking outside the box on who is affected/who has stake

Values to consider:

- Data privacy
- Safety
- Security
- Fun
- Easy to use
- Works / is effective
- Accessibility
- Simplicity
- Has good impact

AI IMPACT MATRIX: LET'S THINK ABOUT IMPACT MATRICES FOR A.I. PROJECTS

	Positives	Negatives	What we'll make	How will we achieve this?
Impact on users/community 1				
Impact 2 on users/community 1				
Impact on users/community 2				

Example 2: IMPACT MATRIX
Problem space: Schools are concerned with integrity of test-taking over Zoom
Possible idea: Monitoring for test-taking while remote?

	Positives	Negatives	What we'll make	How will we achieve this?
Monitoring upholds integrity of test-taking for test-takers (students, direct users) and school admin	People can trust the results of the exam	Lack of trust of students honor code disintegrates student relationships to school. There are also frequently near to "get around it"	Test monitoring over video camera... Should this be the solution?	Let's discuss!
Allows for flexibility of learning/teaching/test-taking remotely for students (direct users)	Students can continue to learn remotely, as well as successfully tests (like SAT) remotely	However, surveillance during test-taking is a "massive privacy" concern.	If using A.I. to process images: online-only inference... what else do you think?	What do you think?
Students and/or school admin may experience non-ethical uses of A.I. (if A.I. is used in the solution)	--	Mistrust of technology, experience of privacy or security violation	Plan it's and other strategies for non-video monitoring of test-taking	Students can request in-person and be accommodated

Example 3: IMPACT MATRIX
Problem space: The elderly and children experience loneliness and isolation at home
Possible idea: Alexa/Google Home/Jibo (social robot) engages them in conversations

	Positives	Negatives	What we'll make	How will we achieve this?
Offers companionship and conversation for elderly and children	Having a conversationalist who is there 24/7, responds, engages	Possibility of no impact on isolation/loneliness because robotic companion rather than human	Lots of conversation topics that continue on in as human a way as possible (instead of dropping the thread)	Make use of large language models that help conversation agents act more human
Keeping up with current events/engaging the mind (for elderly) and learning words/practicing communication (for children)	Engagement in conversation/learning for both elderly and children when no one else is around	Privacy concerns for these especially vulnerable populations	Transparent data policies, default is no data storage to use for offline learning, toggles for opt-in/opt-out	Clear messaging of data use, options for turning mic off, options for saving data for improved services (and removal)
Can take place of parent/guardian/caretaker when they are unavailable	Make emergency calls, address questions, play music, act like robot caretaker when needed	Technology may be unreliable, but may not be clear or transparent for the users that learn to rely on it	Easy & clear options for emergency (perhaps with a 1st), easy options for calling contacts	Possible list for emergency contacts/calling, or practice prompting from agent based on certain triggers

Quantitative Results

Paired Pre/Post: Quantitative data show no significant change in answers to 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.”

Unpaired Pre:

- 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: $\bar{x}=3.172, 2.667$; $p=0.046$; $t(100)=2.02$.)
- Students from Lebanon more strongly agreed with having concerns about A.I. than students from the U.S. (US/Lebanon: $\bar{x}=2.73, 3.6$; $p=0.039$; $U(69)=178.5$)

Qualitative Results

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%)

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 from the perspective of a technology user.

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.