Catalyzing Sociotechnical Thinking: Exploring Engineering Students' Changing Perception of Racism in Automation during a First-Year Computation Course

Dr. Kaylla Cantilina, Tufts University

Kaylla is a Postdoctoral Scholar at Tufts University where her work is motivated by design as a means for social justice. Her research explores the ways that students and practitioners seek to achieve equity in their design practices.

Dr. Ethan E. Danahy, Tufts University

Dr. Ethan Danahy is a Research Associate Professor at the Center for Engineering Education and Outreach (CEEO) with secondary appointment in the Department of Computer Science within the School of Engineering at Tufts University. Having received his graduate degrees in Computer Science and Electrical Engineering from Tufts University, he continues research in the design, implementation, and evaluation of different educational technologies. With particular attention to engaging students in the STEAM content areas, he focuses his investigations on enhancing creativity and innovation, supporting better documentation, and encouraging collaborative learning.

Catalyzing Sociotechnical Thinking: Exploring Engineering Students' Changing Perception of Racism in Automation during a First-Year Computation Course

Abstract

This Complete Evidence-based Practice paper describes first-year engineering students' perceptions, and specifically their shifts in those perspectives, towards the role of automation and data science in society as well as the racial implications of how those human-made systems are implemented and deployed. As part of a larger curricular change being made to a first-year engineering course in computation, this paper specifically examines two reflection assignments where students wrote, at different points in the semester (week 2 and week 12), regarding their personal questions and understandings related to the role of machine learning, artificial intelligence, and automation in society and its relationship to systemic racism and racial impact of engineering and technological systems. For analysis, the submissions were compiled, and comparisons of the two moments in the semester were coded and analyzed for thematic commonalities seen in student written responses and the overall progression of students' thinking. Results showed commonalities among students' initial reactions to the video such as questions surrounding who is responsible for the impact of designed technologies along with a strong ideological separation between humans and machines. Juxtaposed with the week 2 assignments, week 12 findings showed commonalities in students' progress such as an increased awareness of the complexity of racialized sociotechnical problems, stronger emotional responses, more refined ideas about potential solutions, and realizing the systemic nature of racism. Findings suggest that the students met learning goals regarding an awareness of sociotechnical problems and catalyzed (early) critical thinking on how to address them through engineering. Implications from this work demonstrate that first-year students are capable of wrestling with difficult topics such as racism in technology, while still meeting ABET requirements within the course for data science and coding.

Introduction

At a small private engineering institution in the northeast region of the United States, year one of a research-based reimagining and redesign of the "*Introduction to Computing for Engineering*" course was underway in Spring 2022. Imagining a future where engineering students develop, analyze, produce, and assess engineering processes, products, and impacts through a lens of justice and equity, a research and development team was starting this process by integrating social justice topics alongside technical knowledge within this first-year engineering computation course. Across all five sections (taught individually by instructors to class sizes of ~30 to 45 students each) there were three main modifications to the course and course structure being made. First was new content development, with a specific focus on engineering activities that include both computational challenges as well as social and political context and implications. Second, was a shift in course structure to include more reading, reflecting, and in-class small-group discussions around sociotechnical problems, many taken from current event headlines facing our society today. Third, in support of both the instructor and students in the course, the introduction of Equity Learning Assistants (ELAs) to help lead discussions and probe deeper into the thinking, perspectives, and understanding of the students toward the ultimate role that computation (and engineers developing computation-based systems) has on society.

The second change (reading, reflecting, and in-class discussions) was coined "*Computing in the World (CW)*" assignments during this redesign iteration and happened weekly around a wide variety of topics. Below is the sequence topics by week (each paired with appropriate reading, listening, or watching materials), that were used during the Spring 2022 semester:

- CW Week 1: Discussion Protocols
- **CW Week 2: Are We Automating Racism?**
- CW Week 3: Vaccine Passports
- CW Week 4: Environmental Justice
- *No CW assignment in week 5*
- CW Week 6: Energy Justice
- CW Week 7: Impartial Machines
- CW Week 8: Disability in Design
- CW Week 9: Transit Equity
- *No CW assignment in week 10*
- CW Week 11: Racial Bias in Medical Equipment
- **CW Week 12: Are We Automating Racism?**
- CW Week 13: Datasheets for Datasets

Of note are weeks 2 and 12 (bolded, and the subject of this analysis) where the same content was revisited at two different points in the semester and students were asked, in the later moment, to reflect on the material, their previous reflection statements, as well as their shift in perspectives through the semester.

To note is the state of higher education during the 2021-2022 school year: first-year engineering students had just entered university following a high school education severely disrupted by the COVID-19 pandemic. While classes were back in person and social-distancing restrictions had been reduced, masking was still prevalent, routine testing was ongoing, and periodic spikes in COVID cases (and associated on-campus quarantine, isolation, etc) required hybrid class options. Further, students were acutely aware of and discussed events such as the death of George Floyd and the subsequent series of protests, changes to university policies and structures, and other on-campus reactions around diversity, equity, and inclusion. Overall, the collection of "*Computing in the World (CW)*" topics, as well as the development of other computing projects/assignments and the classroom support from the Equity Learning Assistants (ELAs), were all specifically designed to include these ideas and topics into the engineering classroom and show the relationships amongst the skills, tools, and technologies being discussed and the bigger societal impact.

Assignments

The "*Computing in the World (CW)*" assignments for both week 2 and week 12 leveraged the same video content students needed to view. The video entitled "*Are We Automating Racism?*", created by Vox (owned by Vox Media and published on YouTube on March 31st, 2021), challenges the perceptions of tech-as-neutral, highlights the role of data-driven automated systems in our lives, and explains ways in which algorithms can be fundamentally biased. Narrated by host Joss Fong it presents a variety of real-world computing and technology examples (Twitter image cropping of faces, automated soap dispensers, teleconferencing features like face following and background blurring, crime prediction tools based on racial profiling in datasets, health care systems predicting high-risk patients) where automated systems built on

machine learning algorithms and trained on human data exhibit racist behaviors. Through interviews with researchers and experts (such as Dr. Safiya Noble, Deborah Raji, and Dr. Ruha Benjamin among others), the multitude of examples clearly show the problems are not isolated incidents. They also demonstrate a "pattern of harm that disproportionately falls on vulnerable people/people of color," as stated by Dr. Safiya Noble. The video also touches on themes of difficulties of tracking why machine learning models make the decisions they do, the lack of regulatory oversight on machine learning and artificial intelligence, factors of inequities and disparities in resources and power dynamics, as well as leaves open-ended questions around accountability for the viewer to ponder.

Week 2 Assignment Structure

Still early in the semester, before students had become immersed in either the coding content nor the other sociotechnical discussions of the semester, students were asked to watch the video and develop discussion questions they had after viewing the video: *curiosity questions*, *critical questions*, or *application questions*.

- For the video, please come up with 1-2 discussion questions. This can be a curiosity question, where you're interested in finding out more, a critical question, where you challenge the author's assumptions or decisions, or an application question, where you think about how concepts from the reading would apply to a particular context you are interested in exploring.
- In addition to the discussion question, feel free to also ask clarification or definition questions to help us better understand what the author is trying to convey.

Instructions from the Week 2 "*Computing in the World*" Assignment

Week 2 Assignment Structure

Near the end of the semester, after many weeks of both learning Python coding and simultaneously engaging in a sequence of other "*Computing in the World (CW)*" assignments about a wide variety of topics highlighting a wide range of ways in which computing and data science affect the world (often in in-equitable and non-inclusive ways), the students were asked to revisit both the "*Are We Automating Racism?*" video and their previous responses in order to reflect both on their current positionality and how that perspective may have (or may not have) changed since week 2 of the course.

For this week's assignment, I want you to go back and rewatch the video that we saw in Week 02: "Are We Automating Racism?"

I also want you to go back and look at your "Week 02" reflections/questions that you wrote as part of that CW assignment: CW Week 02: Are We Automating Racism?

Think about where we were in Week 02 and what we've done since:

- We had just started the semester and were getting to know each other.
- We were just introducing the first concepts of programming (variables, etc).

- We hadn't done any "Computing in the World" readings, reflections, or discussions yet.
- Since then we also did bigger assignments like the "Solar Panel Project" or the "MBTA/Transit Project".
- We've done a lot of data manipulation, analysis, visualization, and interpretation.

For this assignment, please reflect on:

- How do you NOW feel about the "*Are We Automating Racism?*" video?
- Do you have remaining questions about the content?
- Thinking back to "you in Week 02," are your perspectives the same or different? In what ways?

Instructions from the Week 12 "*Computing in the World*" Assignment

Methods

Data Collection

Responses from the 43 students (of 44 total enrollment) who consented to IRB-approved research were anonymized and collated by students for both assignments. All 43 students submitted a written response to the week 2 prompt and 41 (of 43) students submitted a written response to the week 12 prompt. After reading and reviewing all responses, the researchers developed a set of codes (iteratively refining as needed) to capture trends seen within the data. Each researcher then independently reviewed submissions and applied the coding scheme, and any disparities and variations in coding were discussed till consensus and resolution were reached. Below in Table 1 (cookbook for week 2 responses) and Table 2 (codebook for week 12 responses), are the codes (and explanations) developed for each assignment.

Table 1. Codebook for Week 2 Reflection Responses

Week #	Code Name	Definition/Criteria	Sub Code
2	Technology	Questions that center technology, which includes understanding functionality, manipulation or fixing technology, or proposing technical solutions	
2	Responsibility	Whose fault is it or who should fix it; including if anyone's fault; any discussion of fault. Can include actual groups e.g. companies, policy makers, engineers/programmers	(1) Fault assigned
			(2) No one's fault
2	Society	Thinking about the larger/broader implications of the technology on society, or society's relationship to power and people (e.g. social equity or a particular phenomenon, like racism)	
2	Solutions	Proposing solutions, or stating a solution is not possible, or asking about how to solve it.	(1) can we fix it (via proposal)
			(2) can we not fix it (not possible)
			(3) how can we fix it (open ended)

Table 2. Codebook for Week 12 Reflection Responses

Week #	Code Name	Definition/Criteria	Sub Cde
12	Coding led to improved sociotechnical understanding	Students explicitly saying that understanding or gaining technical coding skills led to a better understanding of the sociotechnical/equity problem	
12	Beliefs have Changed	Students felt their beliefs changed over the semester; they are in a different place now compared to beginning	(1) Changed
			(2) Not Changed
12	Questions Answered	Previous questions from earlier in the semester were answered or not answered	(1) Answered
			(2) Not Answered
12	Level of Pessimism	Expressing it's worse than expected/society is a mess/we can't fix it; or if hopeful and believing "it can be fixed"	(1) Doom: can not be fixed
			(2) Workable: make progress?

Analysis Methodology

We (the two authors) approached the analysis process by reading the responses several times, each time taking notes on different themes or connections we observed, then collectively developed a codebook for each week. Because our sample size is relatively small and our focus is on nuance versus statistical significance, we chose a qualitative thematic approach to make useful observations and discover trends in the data. It is important to note that one key limitation of this work is that the week 2 and week 12 reflection questions were different, thus the connections we draw in the findings are not intended to show direct comparison.

Findings

We present the findings based on our analysis in three sections. The first (*Findings Section 1*) is an analysis characterizing the student reflection responses in week 2. The second (*Findings Section 2*) is an analysis characterizing the responses from week 12. Finally, the third (*Findings Section 3*) highlights key similarities and differences we identified by looking across responses from the weeks.

Findings Section 1: Analysis of Week 2

In week 2, students were asked to form questions that watching the video made them consider. Figure 1. shows the frequency of codes assigned to students' written questions or responses that touched on each identified theme.

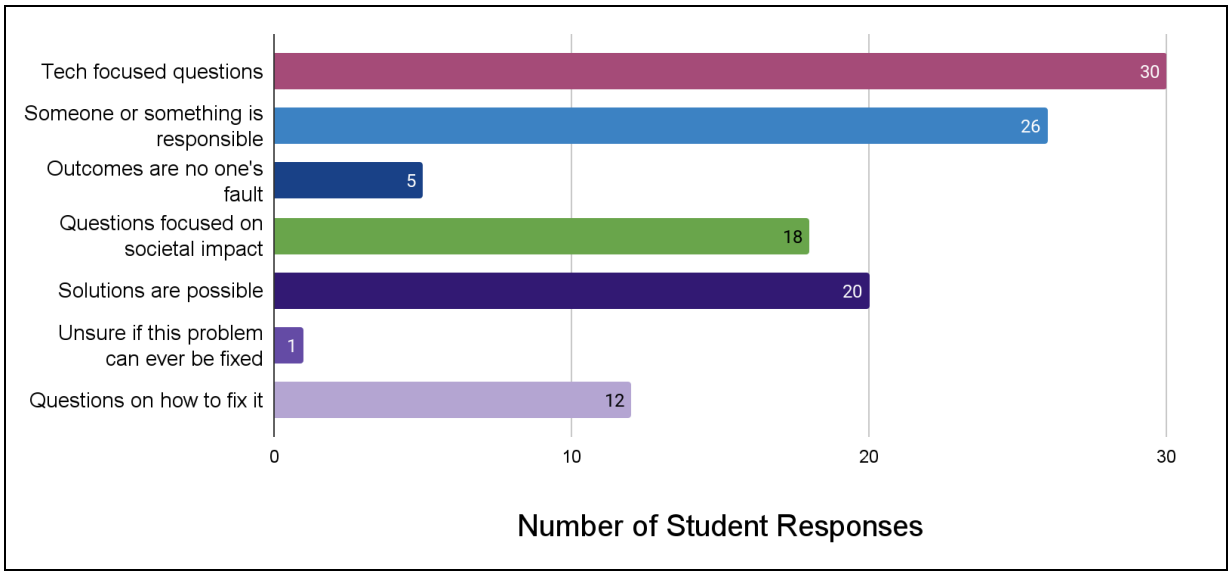

Figure 1. Number of Student Responses Assigned with Each Code in Week 2

The first observation that Figure 1. demonstrates is most students' questions *focused on the technology*, whether it was how algorithms worked, in what way is it flawed, how they could be changed or redesigned to reduce bias, or what technical attributes led to racist outcomes. For example, Student 22 asks:

> *"When it comes to facial recognition, is it coded maliciously? Or is it just created poorly so that white faces are often recognized better than those of other races?" - Student 22*

This is in contrast with questions that focused on the social impact that saw technology as a facilitator, solution, or perpetrator of systemic issues prevalent in society. Responses ranged from questions about the relationship technology has on minoritized communities, to questions about why the intersection of racism and technology was not a bigger part of society's popular discourse. Student 16 captures this sentiment in their question:

> *"Why is it when people talk about Systemic Racism, unjust law, employment and education inequality, stereotypes, and more are mentioned but technological bias doesn't enter the conversation often? Especially when technological advancements (smartphones, social media, the internet, etc.) are usually seen as ways to facilitate systemic racism and misinformation." - Student 16*

While 18 out of the 43 students (42%) did ask questions that focused on *social impact* instead of just about the technology, it was slightly surprising that there were not more social impact questions, considering the video was explicitly about racism.

Most students' questions included suggesting that *someone or something is responsible* for the racist outcomes of AI as described in the video. For instance, Student 8 specifically questions how Twitter has been held accountable for their racialized AI photo cropping algorithms, asking:

> *"Has Twitter done anything to try and fix their biased cropping algorithm since this video came out and displayed it? Have they even addressed or acknowledged it?" - Student 8*

Conversely, other students' questions suggested that there was moral ambiguity to technology that was distinctly separate from human bias. Student 4 grapples with this moral ambiguity in their question:

> *"I personally believe that because a program doesn't really have a moral compass, it's hard to really blame anybody for issues like this, and if you can blame somebody it is a whole other issue." - Student 4*

Interestingly, all the students who asked questions that were coded as the outcome being *no one's fault* also had questions that were *technology-focused*. The students who asked questions suggesting that *someone was responsible* for the racism almost all talked about it in the sense that *those who create the algorithms infuse their own bias into the technology*. For example:

> *"The machine simply does what it's told to do. It does not have an opinion, but a program that is developed by people. In this case, people who developed facial recognition were the ones that are racist, not the software itself." - Student 13*

For the most part, students had hoped that *solutions were possible* to address the racism embedded in AI, though there was still a significant *amount of uncertainty* at this early point in the semester. Student 35 describes uncertainty by asking:

> *"I wonder how much more data will be needed to "balance" the data set. How equal is equal enough? (30/70 was pretty bad, but with other races in the mix, would it even be possible to reach a balance?)" - Student 35*

Findings Section 2: Analysis of Week 12

Towards the end of the semester, the students were tasked with revisiting the same video as in week 2. However, the reflection prompts and overall assignment for week 12 were both very different. As such, the codes used in analyzing the data have to be different as well, which means they do not match one-to-one with the codes from week 2. Therefore, a pre-post comparison approach (that is frequently used to measure course learning or student change) is not possible here. In our analysis, we looked for ways in which students themselves reflect on what they have learned and how they have personally changed from participation in the class. Figure 2 shows the number of appearances of each code (from the week 12 codebook) amongst all student responses submitted in week 12.

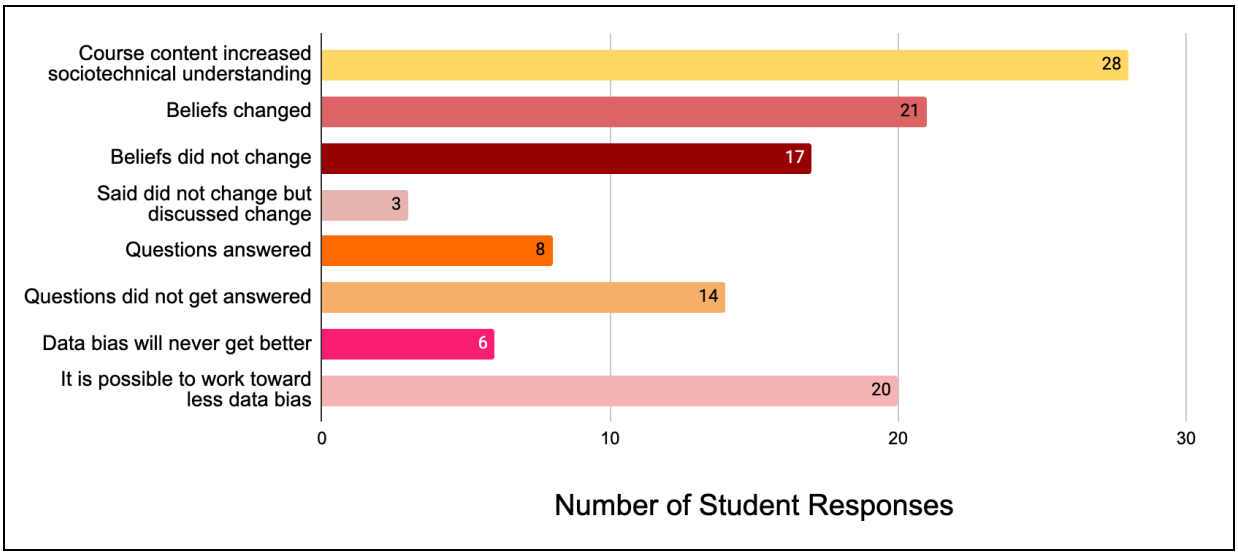

Figure 2. Number of Student Responses Assigned with Each Code in Week 12

The code with the most number of appearances (28 of 41, or 68%) is that a majority of students felt that the *course content (both learning coding skills and the computing in the world labs) led to a greater understanding of the sociotechnical nature of data science*, specifically how injustices or inequities like racism are perpetuated through technology. This student describes how learning led to their better understanding, saying:

> *"Through the weekly assignments and lectures, I've grasped and learned a variety of powerful and versatile programming concepts from functions to variables to loops. On the other hand, the weekly Computing in the World discussions added more layers of depth and nuance into the relevance of the aforementioned programming concepts in the lenses of societal themes from medicine to social justice." - Student 33*

This increased understanding also led many students to experience certain levels of self-actualization and awareness about the inequitable facets of society that are exacerbated by technology that they previously did not have. We noted this as students' *beliefs changed*, as their perspectives were broadened. For example, Student 24 discusses how his positionality led to an initial assumption about the problem, but over the semester his mindset changed:

> *"At first, I was kind of dismissive, as these problems don't personally affect me. After all, I am pretty much the poster child of privilege: I'm white, male, wealthy, Christian, Western European, and have been given so many advantages in life through no merit of my own. These biased algorithms were basically built by people like me, for people like me. Now, I have learned about these algorithms, how they affect people, and how we can work to fix them." - Student 24*

Most of the students who reported that their *beliefs remained the same* as the beginning of the semester already held beliefs in week 2 that aligned with the main intended takeaways from the video, such as that data, algorithms, and machine learning systems can replicate human bias, leading to inequalities. Student 30 explains their unchanging beliefs saying:

> *"I feel the same as I did at the beginning when I first watched it. AI and technology are biased generally towards white men. That's who writes the programs…" - Student 30*

Other students who expressed that their beliefs or perspectives had not changed often highlighted that despite this, the class helped them understand their beliefs better or provided more evidence for why they believe what they do. This does lead to some implications that the course benefits a wide range of students, regardless of their level of awareness, belief, or value alignment with the specific course content.

Many students felt empowered by gaining knowledge from taking the class. Having now been equipped with new tools and skills to problem solve and understand the problem space, students expressed that they could begin to come up with actual *ideas to address data bias and racism in machine learning*. Student 17 describes their experience and new mindset saying:

> *"These problems have already been created so instead of focusing on ways that could've been more beneficial in the past, we should focus on what we can do now to turn these issues around. Maybe even code an algorithm [sic] to detect these issues ahead of time." - Student 17*

However, not all students felt equally empowered; rather some felt that a better understanding of it had actually added to a growing sense of feeling overwhelmed and that the prevalence of equity issues tied to data is a significant problem with no true solution. Student 25 describes how their gained knowledge led to more confusion about a path forward:

> *"First watching this video, I had the idea in my head that this was a fairly simple problem to solve. I thought that though it may be tedious, the concept of fixing this problem was simple - add more representation to the data. Now that I realize how complicated even the basics of coding are, going back and changing all data to be fully inclusive seems like a basically impossible task." - Student 25*

Lastly, for students who expressed their questions remained unanswered, there was not a noticeable pattern in the types of questions. However, not all students confirmed (or directly or explicitly addressed) in their reflections whether their questions had been answered or not, so this dataset is incomplete.

Findings Section 3: Looking Across the Two Weeks

There are a few noteworthy observations we made when looking at the data from the two weeks side-by-side. The first is that overall, we saw a shift from students' early thinking about racism in AI being a strictly technology problem that could be solved by fixing the technology itself, to week 12 where the majority of students came to the understanding that social dilemmas were indeed impossible to disentangle from the technological context. This shift was also noticeable when looking at what original questions students felt were answered. The questions that students did feel were answered through the course of the semester had to do with *how the technology worked* and the *process through which algorithms were reproducing racism*. Many of the questions that did remain were non-technology oriented, such as why policies have not better-addressed bias in data and technology, or who society should be holding responsible for the issues.

Aside from ideological shifts, we also saw shifts in the language students used to talk about the topic (while acknowledging the difference in assignment structures). In week 12, students shared their emotional reactions, current feelings, and personal relationships to the problem. This is

compared to the week 2 responses which were primarily depersonalized and intellectual in nature. There was also a stronger sense of urgency to address the problem by the end of the semester. In week 2, students discussed the racism in machine learning problem with curiosity, while in week 12, students felt the problem should be addressed immediately and with more emphasis from different parts of society whether it be engineers, business owners, the government, etc. While this could be an outcome of the difference in assignment prompts, we still believe that these differences were observable and meaningful, especially because the students do reflectively discuss the changes they have personally gone through explicitly. Ultimately, within the context of increased knowledge, issue urgency, and personal connection to the problem described in the video, the majority of students still expressed hope that addressing and even solving the problem was possible.

Discussion & Conclusion

First-year intro engineering courses are typically purely technical in nature, and as such students can bring preconceptions, and have those developed and reinforced, around the ideological separation between humans and machines. Evidence from the assignment early in the semester (week 2) showed commonalities amongst students' initial reactions to the video that confirmed this perspective: students focusing on the technological aspects of the problem, questioning who (on an individual level) is responsible for the impact of designed technologies, and communicating in their responses a sociotechnical divide.

Juxtaposed with the week 2 assignments and after a semester of not only engaging in the technical content of learning to code but also sociotechnical projects and weekly "*Computing in the World (CW)*" readings, reflections, and discussions, the week 12 responses showed increased refinement of ideas and more nuanced understandings of the societal impacts of engineering. Findings from the later reflections showed commonalities in students' developmental progress such as an increased awareness of the complexity of racialized sociotechnical problems, stronger personal and emotional responses, more clarity about potential solutions (or complexities related to solutions), and deeper realization of the systemic nature of bias in algorithms and data.

Responses from the end of the semester showed students still having several remaining questions, concerns, and in some cases anxiety about the path forward towards developing universal solutions to these systemic problems. Because these are just first-year students, this is viewed as a positive, as the hope is they carry these uncertainties with them to their subsequent engineering classes and continue to develop and refine their understandings of the sociotechnical interplays amongst engineers/developers, the tools, and technologies, and society.

This work demonstrated the possibility of incorporating sociotechnical topics into first-year courses alongside and integrated within technical content. Implications from this work demonstrate that first-year students are capable of wrestling with difficult topics such as racism in technology, while still meeting ABET requirements for data science and coding. Students' perceptions and perspectives of data bias and racist algorithmic impacts not only grew in depth and refinement, but students were self-aware and actively engaged in reflecting on that process. This not only implies a need for this type of content in the first year of engineering but the larger need for successive courses as well throughout the engineering curriculum where students are provided ongoing opportunities to further develop their perceptions and understandings of these topics.