7-26-2021

Raising Algorithm Bias Awareness Among Computer Science Students Through Library and Computer Science Instruction

Shalini Ramachandran
Boise State University

Steven Matthew Cutchin
Boise State University

Sheree Fu
California State University, Los Angeles

2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #34964

Raising Algorithm Bias Awareness Among Computer Science Students Through Library and Computer Science Instruction

Shalini Ramachandran, Boise State University

Shalini Ramachandran is a Faculty Liaison for Research Development at Boise State University. Prior to this position, she was a Science and Engineering Librarian at the University of Southern California. Her research interests include algorithm bias, information access in higher education, and open access publishing.

Dr. Steven Matthew Cutchin, Boise State University

Dr. Steve Cutchin joined the faculty at Boise State University in August 2013 From 2008 to 2013 he was manager of the KAUST Visualization Laboratory Core Facility and the Supercomputer Facility at King Abdullah's University of Science and Technology (KAUST) in Thuwal, Saudi Arabia. At KAUST he recruited a technical team of engineers and visualization scientists while managing the building of the state of the art scientific data visualization laboratory on the KAUST campus, forged relationships with international university and corporate partners, continued to improve the laboratory and recruit new staff. Prior to his work in Saudi Arabia, Dr. Cutchin worked at the University of California, San Diego (UCSD) first as manager of Visualization Services at the San Diego Supercomputer Center and later at California Institute for Telecommunications and Information Technology (Calit2). He has worked as a Sr. Software Engineer at Walt Disney Feature Animation developing software tools to improve animation production on feature films. He has published articles on Computer Graphics and Visualization, created animations for Discovery Channel and images for SIGGRAPH and Supercomputing conferences and journals. He received his doctorate from Purdue University in Computer Science.

Sheree Fu, California State University, Los Angeles

Sheree Fu is the Engineering, Computer Science, and Technology Librarian at California State University, Los Angeles.

Raising Algorithm Bias Awareness among Computer Science Students through Library and Computer Science Instruction

Abstract

We are a computer science professor and two librarians who work closely with computer science students. In this paper, we outline the development of an introductory algorithm bias instruction session. As part of our lesson development, we analyzed the results of a survey we conducted of computer science students at three universities on their perceptions about search-engine and big-data algorithms. We examined whether an information literacy component focused on algorithmic bias was beneficial to offer to students in the computational sciences and designed an instructional prototype. We studied qualitative data, including feedback from students and colleagues on our initial instruction module to create the next two modules. We found that students' reception to the subject of algorithm bias can range from defensive and unaccepting to open and accepting of the existence of such bias. Since the topic ultimately deals with issues of racial, gender-based, and other discrimination, a multidisciplinary approach is needed when teaching about algorithm bias. Our assertion is that librarians have a role in partnering with computer science instructors to ensure that students who major in computer science, who will be the primary creators of algorithms as they enter the workforce, can develop an early awareness and understanding of bias in information systems. Further, when the students receive such training, the automated systems they generate will produce more fair outcomes. Our pedagogy incorporates insights from computer science, library science, medical ethics, and critical theory. The aim of our algorithm bias instruction is to help computer science students recognize and mitigate the systematic marginalization of groups within the current technological environment.

Introduction

Search-engine bias and unfair outcomes from automated systems have been documented in recent years. All modern information systems depend on computer algorithms to run effective programs. Algorithms are sets of instructions within computer programs that direct how these programs read, collect, process, and analyze data. We use the term bias to refer to computer algorithms that systematically discriminate against certain content, individuals, or groups without a sound basis [1].

As automated systems become an integral part of many decisions that affect our daily life, civil rights, and public discourse, there is concern among social scientists and computer scientists about the presence of bias in machine learning and big-data algorithms. A body of work has appeared in popular as well as scholarly literature addressing algorithm bias. In 2018, then visiting assistant professor at the University of Southern California, Safiya Noble

[2], who also holds a faculty appointment at The University of California, Los Angeles (UCLA), published a book, *Algorithms of Oppression*, where she details significant bias against women and people of color within the Google search structure.

Many practicing mathematicians and computer scientists have also tackled the issue of bias and fairness in mathematical algorithms. Cathy O'Neil [3] points out in her book, *Weapons of Math Destruction*, that human bias can be encoded into mathematical models. She calls the faulty predictive models "weapons of math destruction" (WMDs). An example she gives is of racism as a poorly designed mathematical model: "Racism, at the individual level, can be seen as a predictive model whirring away in billions of human minds around the world. It is built from faulty, incomplete, or generalized data. Whether it comes from experience or hearsay, the data indicates that certain types of people have behaved badly. That generates a binary prediction that all people of that race will behave that same way." O'Neil is not arguing that computer algorithms are created by racists. However, she suggests that the models powering artificial intelligence (AI) and other computational tools mimics some of the aspects of how racism operates: "Racism is powered by haphazard data gathering and spurious correlations, reinforced by institutional inequities, and polluted by confirmation bias."

The academic and professional contours of computer science have changed dramatically in the past 20 years. Early computer science programs were housed in schools of science, as the discipline grew out of mathematics and applied mathematics. Later, as computer software and hardware decision making became more intermeshed and the computer industry became a significant part of the technology sector, many institutions began to locate their computer science departments within schools of engineering. At all three of our home institutions, the computer science department is within schools of engineering. For the two engineering librarians in this study, computer science is one of our subject responsibilities. As such, for our study, we have included computer science as an engineering discipline. We understand there is some discussion of where computer science fits as a STEM discipline, but that is outside the scope of our paper. Being instructors and librarians in institutions where computer science plays a prominent role in engineering schools, we address computer science education as an integral sphere of engineering education. Further, while our focus is on algorithm bias as a segment of computer science ethics education, many of the core ethical values discussed herein can be extended across engineering fields.

From a professional perspective, the sector, now called the "tech industry," has transformed as well. In the past, a computer science graduate would expect to work on in-house systems at a technology firm. These days, even before they graduate, computer science students work as interns in public facing web systems with the potential to be deployed to millions of people. Given that level of penetration that computer algorithms have into people's private decision making processes, training our undergraduate students in ethics pertaining to algorithms

becomes a necessity, much like medical ethics are required of medical students. Computer systems may have bias in their operation. Our students need to have an understanding of how bias enters these sophisticated software applications and how to prevent it.

We researched algorithm bias education of computer science students because we wanted to develop a module on bias awareness and assessment for our students. Our aim is to help students in their college coursework, and, later, as practicing computer programmers, to create software systems that produce fair outcomes for individuals in society.

Harms of Algorithm Bias

There are many reasons why an algorithm may be considered "biased." Incomplete or faulty data is one reason. In other instances, it may be the choice of data that is being selected for decision making. As an example of the latter, in a 2019 paper [4] in *Science*, "Dissecting racial bias in an algorithm used to manage the health of populations," authors Obermeyer et al. found evidence that a widely used commercial prediction algorithm for determining health risk was consistently scoring black patients as being lower risk for health issues than white patients, even though, in their study, the black patients had significant health risks that were being missed by the algorithm. As they point out, "The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for white patients. Thus, despite health care cost appearing to be an effective proxy for health by some measures of predictive accuracy, large racial biases arise." Thus, in this case, the emphasis on health care cost data in the predictive algorithm overlooked the ground reality of unequal health care access, which meant that black patients were not receiving as much medical care as white patients, and thereby making it appear that they were healthier than they were because of a lower health care expenditure.

Another reason for algorithm bias is the possibility of bias inserted by humans. For instance, Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) is a case management tool widely used in the US to guide sentencing by predicting the likelihood of a criminal reoffending. In May 2016, ProPublica [5] reported that the COMPAS system predicts that black defendants pose a higher risk of recidivism than they do, and the reverse for white defendants. Equivant, the company that developed the software, disputes that. It is hard to pinpoint where this bias might come from, because the algorithm is proprietary.

We point to two recent examples of algorithm bias, but there are many more in existence as a quick search of current news sources will reveal. What is significant is that even with the two examples we identify, millions of lives are potentially affected through algorithms that are making automated choices. Because of the wide-reaching power of algorithms to impact our public and private spheres, it becomes crucial that we intervene early and often within the

computer science education of students to familiarize them with the phenomenon of algorithm bias.

Methodology

We conducted our investigation primarily through case studies at three institutions. Our objective for the instruction module development process was three-fold:

1) To examine existing computer ethics education currently taught to undergraduate students to get an idea of how algorithm bias is discussed in computer science education.
2) To understand computer students' perceptions of algorithm bias.
3) To develop an instructional prototype where the algorithm bias is introduced to computer science majors early in their university education.

We based our work in three institutions, Boise State University in Boise, Idaho; University of Southern California in Los Angeles, California; and California State University in Los Angeles, California. The initial phase of the study involved a review of the state of ethics education in computer science. We examined literature and looked at course offerings at our institutions and others to get an idea of how ethics and, more specifically, algorithm bias is taught in computer science programs. Next, we conducted a survey of computer science students at all three institutions and based on results of the survey, developed an introductory instructional module which was first tested as a guest lecture in an existing Computer Science Special Topics class in Spring 2019. The lecture was revised into a module and was taught again at Boise State University on July 24, 2019, at University of Southern California on November 25, 2019, and at California State University, Los Angeles on March 11, 2020.

1) State of Ethics Education in Computer Science

Our literature review explored published work in two broad areas: a) pertaining to how ethical concerns have traditionally been addressed in computer science and engineering education, and b) pertaining to how the phenomenon of algorithm bias has been addressed by instructors in computer science and engineering. Additionally, we also searched for current newspaper and magazine articles on our topics of interest.

a) Ethics Education in Computer Science

Some of the literature we found on ethics education did not focus exclusively on computer science but more broadly on engineering ethics as a whole, which we chose to include in our discussion because core ethical decision making processes can be similar across the engineering disciplines of which computer science is a part. Among literature of note was

Athey [6] which surveyed a collection of undergraduate computer science and computer information systems students to determine if they agree or disagree with the ethical determination by identified ethics experts within the fields on their assessment of scenarios and ethical problems. The students notably disagreed with the trained experts in half of the identified scenarios. The disagreement between students and experts possibly shows that further exposure to real-world scenarios may be needed in engineering ethics courses. These kinds of disagreement may have implications for algorithm bias instruction as well because students are not trained to respond to cases of systemic bias but may be responding to scenarios from an individual perspective.

A paper by Bowers, Maccarone, & Ricco [7] discussed their experiences integrating ethical, legal and societal issues within a senior design computing capstone program. The course integrated consideration of an established code of ethics into a capstone program that involved undergraduates implementing computational solutions. It was observed that students' concern for ethics in assessment tools increased at the close of the course. It was not discussed if students' concern for ethics were derived from the construction of their computational system or from the study time spent on topics of ethics within the course. Again, our hypothesis is that exposure to ethical concepts can have an impact in how students approach the construction of an algorithm regardless of whether the change is because of "study time" or from gaining new knowledge. The two aspects work synergistically, and, therefore, we support continuous presentations of ethics concepts to students in all coursework rather than concentrate computer ethics into a one semester requirement.

Another paper of note was Hilliger, Strello, Castro, F., & Pérez-Sanagustín [8], which implemented a quantitative assessment tool for measuring the effectiveness of an ethics course on the thinking of engineering students at a selective engineering school in Chile. The tool identified differences in ethical reasoning between student subgroups along gender and socio-economic lines. The authors concluded these factors impact ethical reasoning even in light of ethical training. This conclusion is significant and corroborated our own observations that students from different racial backgrounds or those who had more exposure to varyingly diverse environments responded differently to the concept of algorithm bias.

A final article we examined was Hedayati Mehdiabadi [9] which reviewed data collected from 33 undergraduate computing majors to learn about their ethics making process. It was determined that engineering students' decisions are highly influenced by the situation and the nature of their decision-making can be regularly ad-hoc. It provides evidence that students lack training on a defined process for ethical decision making. This evidence was also significant in that it underscores the need for a more comprehensively developed ethics education for our computer science students faced with large scale problems that require systematic evaluation rather than ad hoc decision making.

Typically, the papers we studied relied heavily on the utilization of case studies with an underpinning of readings. Assessment of ethics courses relied on surveys, interviews, and tests without any long term summative studies seen to reassess later in a student's educational process. Multiple papers [10]- [11] specifically cited the Accreditation Board for Engineering and Technology, Inc (ABET) compliance as a strong motivator for the creation of the curriculum.

We also looked at course materials used at one of our home institutions, Boise State University, which utilizes texts and resources [12]-[13] for teaching undergraduate students about their ethical responsibilities within the field of computer science as part of a required course for computer science majors. One of the textbooks used is *Ethics for the Information Age* (2017) by Michael Quinn [12]. Boise State University course materials provide a strong grounding in fundamental ethical issues and topics such as intellectual property, patent, and copyright. However, software development has rapidly changed within the past few years and computer scientists, today, work on systems that will be utilized by potentially hundreds of millions of users on a daily basis. The accelerated pace and reach of this new workplace introduces new ethical considerations which require additional models of training in algorithm design for computer science majors.

b) Algorithm Bias Education

Algorithm bias instruction is an evolving field. When we first embarked on a review of literature that specifically looked at algorithm bias instruction, we did not find recent published work in the field. However, compared to when we began our research in 2018, we find that many computer science programs now recognize and incorporate courses on algorithm bias into the curriculum. An article published in 2019, "Embedded EthiCS: Integrating Ethics Across CS Education" by Grosz et al. [14] outlines efforts at Harvard University to create a multidisciplinary approach to teaching ethics to CS students, utilizing instructors from philosophy and computer science to teach courses that address various ethical issues that face computer scientists.

The Embedded EthiCS approach is being embraced at other institutions as well. Stanford University [15] is one of the universities taking a lead in introducing several courses that address algorithm bias, among them CS 384, Ethical and Social Issues in Natural Language Processing, where students study how gender and racial bias can be perpetuated by algorithms and how language processing can be used in a way to address social problems. The computer science department at Stanford is within their School of Engineering. CS 384 is taught by Dan Jurafsky, who holds dual appointments in computer science and linguistics. Several such courses are burgeoning in universities throughout the country. To be noted is that many of the

courses are interdisciplinary, bridging the humanities, social sciences, medicine, and computer science.

These disciplinary crossovers are encouraging and necessary. In the future, we would also like to see metacognitive scholarship on how algorithm bias is taught, evaluated, and assessed for success, as well as the introduction of the topic in earlier points of undergraduate study.

2) Survey to Understand Computer Science Students' Perceptions of Search Engine and "Big-Data" Algorithms

Our survey of the state of ethics and algorithm bias instruction helped us identify gaps that our instruction could address. However, before we set out on developing our preliminary module for algorithm bias instruction, we wanted to know how much students of computer science perceived bias in algorithms and what their thoughts were on how bias could be mitigated. We conducted a survey, which we later published [16], that was deployed to Boise State University, University of Southern California, and California State University, Los Angeles computer science students. Our survey included undergraduate and graduate students.

Although surveys on algorithm bias have been conducted in the past, notably, by Pew Research [17], the computer science student population has not been specifically studied. The significance of our survey is that it studied a population that would be directly participating in the analysis, coding, and design of future computer architectures.

Survey Results

From December 3, 2018, until March 11, 2019, we surveyed computer science students at all three universities. The participants were recruited broadly via listservs and newsletters and included responses from a range of computer science students from first year to graduate students. It should be mentioned that our study is not intended to be a complete formal quantitative investigation. Validation of the results with larger studies may be required.

The total number of raw data responses from all three institutions was 815. After cleaning the raw data to remove responses without signed consent, the total number of responses was 782. The full set of questions that were asked is included in Appendix A.

Opinions of the respondents regarding the questions on search engine results and algorithm bias were recorded in the form of a 7-point Likert scale ranging from "Strongly disagree" to "Strongly agree". A sampling issue with the respondents was that graduate students are majority international students (85% for master's) while undergrad students are mostly U.S.

residents (88%). This made it uncertain about the underlying cause or effect of degree level or international status on opinions. We made the choice to analyze from the perspective of degree-level rather than citizenship.

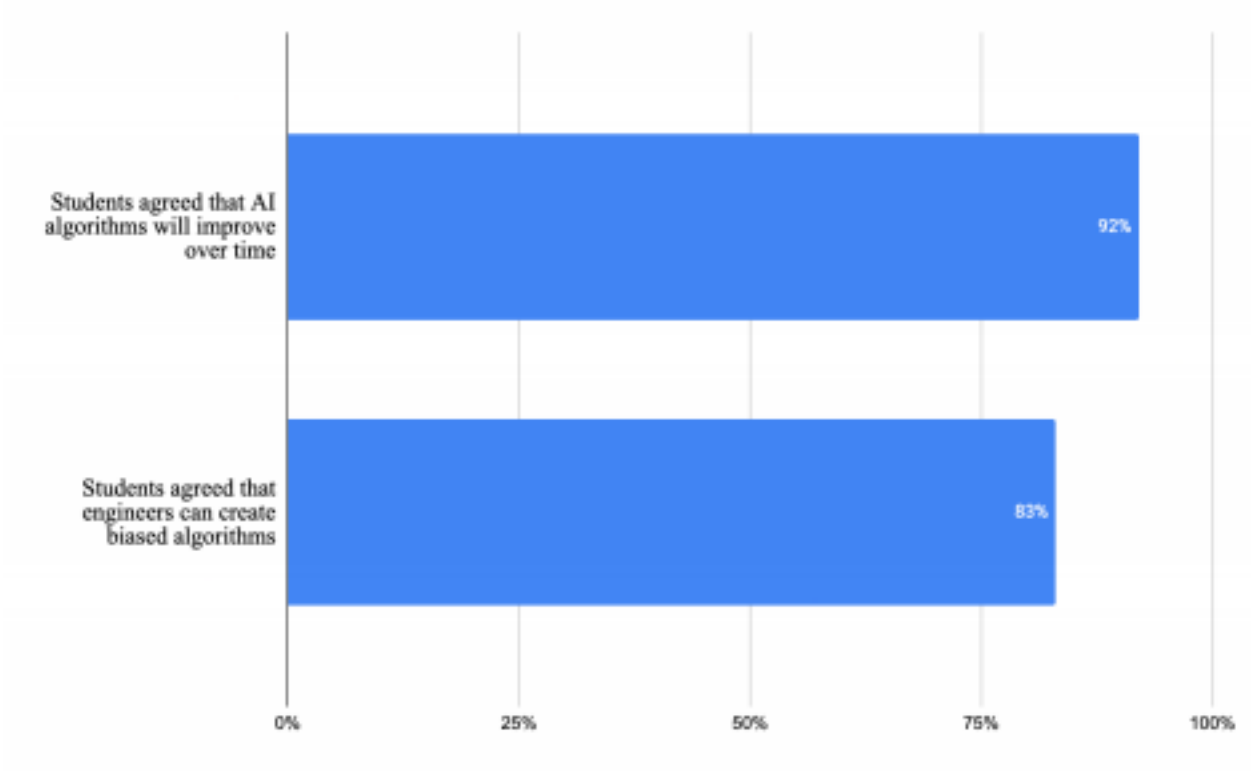

Fig. 1. Student perceptions: AI optimism and biased algorithms

While the general survey findings (Appendix B) informed our overall approach to instruction, what stood out in the survey is that 92% of students agreed that AI algorithms will improve over time. Computer science students present an overwhelming techno-optimism about the future of AI code. A belief that technology can be harnessed to help make technology better was consistently presented. Simultaneously, 83% of our respondents strongly agreed, agreed, or somewhat agreed that “engineers can create biased algorithms.” These two responses in combination would seem to indicate the possibility that those being trained to develop AI code recognize the possibility of bias but hold an unwavering optimism that could reasonably lead them to overlook this potential in their professional work. This means that a systematic introduction to bias in computer systems would be beneficial for students.

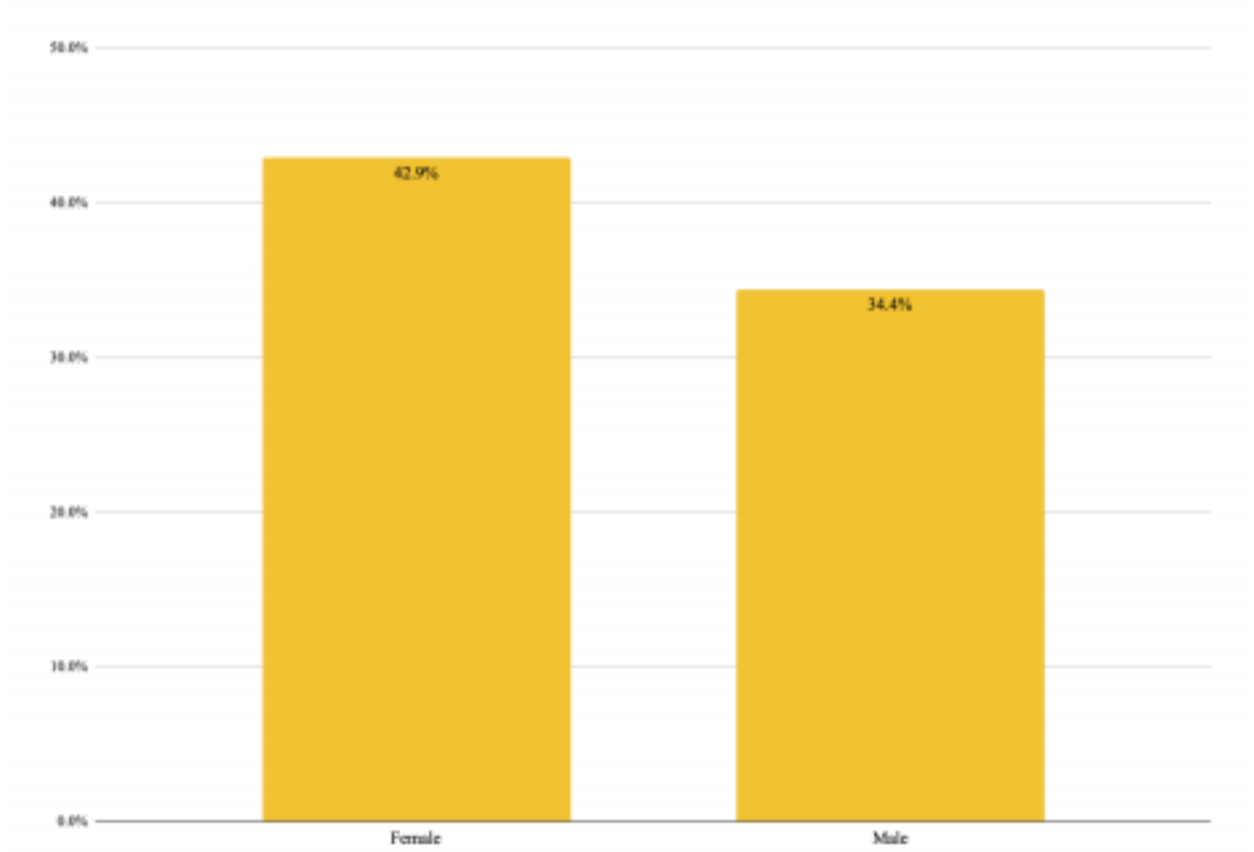

Fig. 2. Students' perceptions: gender stereotypes in search results

We asked students if they encountered bias as they searched the internet. Their response to "I have unexpectedly seen search results that reflect gender stereotypes when doing a routine search," was that 42.9% of females report as true compared to 34.4% of males. Furthermore, students seem to recognize gender bias when searching online with Latinx students reporting the highest at 28.8% with the least being those who identified as Other at 16.1%. Our survey did not directly ask students whether they had heard of or were aware of the concept of algorithm. bias. The survey questions were to determine if students had perceived any bias in their daily interaction with search engines and AI.

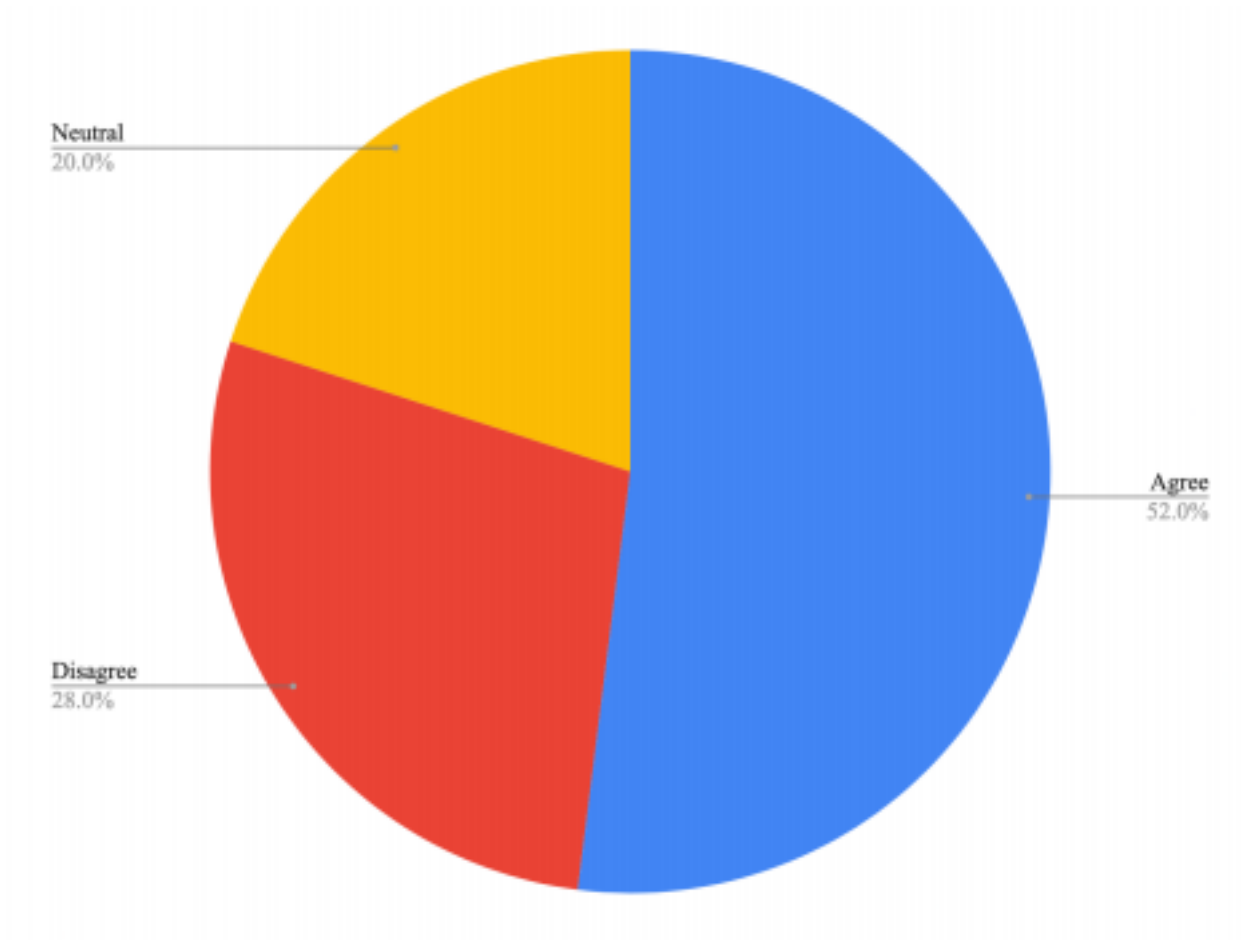

Fig. 3. Government should regulate search engine results

Additionally, most students (83%) agreed that private companies should regulate themselves with regard to correcting systemic bias in their products. However, we did not get consistent agreement from students for government regulation. Although 52% of students agreed that the government should regulate for fair search engine results, the remaining 28% disagree, and 20% remain neutral. As scholars and practitioners trained in information systems, we find that a regulatory environment is necessary to mitigate harm, especially as many of the existing private internet and computer corporations have consolidated power and monopolize the technology market. This aspect of regulation and oversight is not part of the instruction module we have initially developed, but we feel that as algorithm bias instruction takes root in computer science curricula, discussions of the concentrated power of private corporations and antitrust regulation will need to occur.

3) Instruction Module

Boise State University

Based on findings from our survey that students generally perceive algorithms, specifically search engine algorithms as unbiased, we concluded that most students in our study were not aware that algorithm bias exists and can cause social inequity. During the initial guest lecture in Spring 2019 at Boise State University, we also received some push back from students on the existence of algorithm bias as well as the definition of bias. Based on that first class interaction, we developed our instruction with a clearer definition of bias. The design of the class began with creating the following learning objectives:

- Students will be able to define algorithm bias and recognize examples of algorithm bias
- Students will be able to summarize the causes and harms of biased algorithms

The module was launched in late July 2019 for 20 students in a 200-level computer science class at Boise State University. We introduced the topic to students through definitions and examples of bias and algorithm bias. Further, we created a foundation and context for critical thinking and deeper learning by demonstrating examples of which specific algorithms are subject to bias. We spent significant time differentiating between algorithms with a clearly testable result (sorting numbers) and algorithms without testable results (the best pizza in downtown). We presented examples of algorithms that provide recommendations to complex problems without definitive answers rather than examples that had more simplistic (one correct answer) solutions. One example we used was asking the question what the best way is to grow a medium size city.
The next section of the module focused on the challenge that recognizing algorithm bias is difficult but not impossible. Although a lecture with slides was the primary form of delivery, students had opportunities to ask and answer questions during the instruction session.

Material presented to students was guided by an expectation of the orientation of the students in computer science majors. It was expected that students would be highly analytical, critical and questioning of material presented. To this end a set of guidelines for presentation of material was chosen for inclusion. The following criteria were utilized in choosing content:

- All material and content were expected to utilize definitions presented within the material itself and clearly available through standardized reference materials.
- Case study material demonstrating algorithm bias were excluded if it was not possible to identify a specific clear and measurable harm. There were case studies of algorithm bias that demonstrated societal or reputational harm without a clear metric for harm that

were excluded.

- Case studies were also excluded when it was not possible to clearly explain in a manner of construction familiar to the students how the system overall introduced and/or reinforced bias. For example, resume systems frequently introduce bias in candidate selection but we could not locate one that could explain how the bias was being introduced and translated through the system in a clear step by step manner.
- As frequently as possible controversial material involving race or gender were presented in a grounded factual basis where the full facts and totality of a situation could be presented and the direct measurable consequences of algorithm performance could be shown.
- Emphasis was placed on the consequences of algorithms and how unintended consequences could occur and were clearly presented.

With this rubric in place we did remove possible examples from our module when it was not possible to clearly present all of the facts involved.

Preliminary Feedback

The majority of the students were completing a major or minor in computer science. The class began by asking students for their definition of algorithm bias. None of the students knew what it meant. At the end of the class, four students completed the assessment of learning objectives to evaluate any changes in their knowledge of the topic. The assessment consisted of four multiple-choice questions and one short-answer question. Despite the low rate of responses, the responding students recognized a definition of algorithm bias, examples of algorithm bias, and examples of its negative consequences. When students were asked to write a short summary and include their main takeaway, two of the four students mentioned the harms of bias with the following comments:

> "Algorithm bias can harm certain groups of people and it can be mitigated through awareness."
>
> "Information bias is much more prevalent in our lives than we think and it can be extremely dangerous to peoples lively hoods (sic) and freedom. information bias should always be looked for and prevented when creating programs."

Following the instruction, we interviewed two students and two librarians about their thoughts on the instruction. Both students were new to coding and algorithm bias. One student thought the instructor was bringing attention to the problem and appreciated the effort to define it and make it meaningful and more clear with examples. What seemed unclear was how to solve the problem. A possible solution of using more people to review the code and results seemed difficult to implement, according to the student.

The librarians who were interviewed were also not familiar with the topic. They provided comments as new learners and experienced educators. They commented on the topic's timeliness, impact, and relevance. It resonated with them on various levels. One librarian shared her experiences of bias as well as an example of her own inherent bias. One mentioned that technology is great but we need to be careful with it as we are fallible as humans. Both agreed that humans inherently have cognitive biases.

The librarians' pedagogical comments included that the lecture was clear and well-explained. They appreciated the examples and the encouragement such as the instructor saying, "there were no wrong answers." Their suggested improvements centered on active learning and formative assessments. One librarian remarked that for her learning preferences, less text on the slides would have reduced the cognitive load. She was forced to choose between reading the slide or listening to the lecture. The second librarian mentioned that should the module be delivered online, an interactive formative assessment could be included to check if students understood the information from each section. She also suggested that the algorithm bias examples could be customized for different disciplines and subjects. Overall, the feedback was encouraging and constructive. The respondents demonstrated that they recognized definitions of algorithm bias and its harms.

Additional presentations at University of Southern California and California State University, Los Angeles

Based on the work done in this current project we identified areas in the module that needed refinement and adjustment. After making these changes to the initial material the module was deployed in sessions at Boise State University, Institution3, and California State University, Los Angeles in Summer 2019, Fall 2019, and Spring 2020 respectively.

The module identified a grounding definition of algorithm bias that explained the difference between algorithms such as sorting that could not have bias and complex algorithms that could have bias such as facial recognition.

Further, based on earlier student responses in an earlier test presentation specific examples were removed and replaced with new ones that provide a more transparent example of the biases being introduced by the system:

- An initial case study of algorithm bias for resumes used as an example at Boise State University was removed because it did not clearly explain how the bias was occurring within the system.
- Another case study for algorithm bias in search was removed because it did not

clearly demonstrate a measurable harm to the individuals discussed within the study.
- An example of algorithm bias in credit calculations was added because the example involved a married couple whose assets were owned jointly and the only possible explanation was algorithm bias for the generation of different individual credit scores.
- A case study for medical risk assessment was added because it demonstrated the process that led to the algorithm bias and how the bias led to direct harm for the patients.

The introduction of clearly analyzable case studies eliminated defensive questions about the existence of bias in the cases presented. The module with this framework was well received by students at both institutions. The students at University of Southern California were the most familiar with the concept of algorithm and engaged with the material in a similar manner to the students at the other institutions.

Overall observations

The responses at all institutions were similar but with some specific differences:

Boise State University students

- Students would have liked to see more detail.
- Students came away with a better understanding of algorithm bias.
- Tended to argue against the presence of bias when the situation was unclear.

University of Southern California students

- This group was highly familiar with algorithm bias.
- The majority believed that Google search was biased.
- Opposite of Boise State University: often argued for presence of bias when the situation was unclear.

California State University, Los Angeles students (taught remotely due to the COVID-19 pandemic)

- This group appeared to be least familiar with algorithm bias
- The majority believed that Google search was biased.
- Because it was online, it was difficult to gauge student reaction during the presentation.

Drawing on feedback during our presentations, future adjustments to the module are planned. The module will be modified with more specific examples that directly address how algorithm

bias is relevant for computer science students to study. There will also be further discussion of steps practitioners can take to correct algorithm bias in their professional work. Some areas to consider as we further refine the design of instructional modules include:

- Bias in software designers' perspectives can be transferred to the algorithms they write.
- Search engine results may also be biased because the information that the search engines crawl may have gaps (i.e., if the information is not available to be crawled and discovered, it will not be represented in search results; thus there is interest in Wikipedia Edit-A-Thons where users fill in the gaps in information about underrepresented groups such as women, minorities, identities, non-western cultures, etc.)
- This raises the issue of making algorithm design, in the case of search engines, transparent, so users understand why they are getting the results that they are, so they can evaluate the results.

Questions expected to be raised by the engineering and computer science communities within this field are:

- How can we design and implement algorithms that can computationally reason with algorithmic bias?
- How can we explain algorithmic bias in as readily accessible a way as we can explain a 2 degree cutting bias in a table saw?
- What biases do engineers bring to the broader study of algorithmic bias? For example a common attitude amongst computer scientists is the expectation that any problems introduced by new technology will be resolvable by the further development of technology.

Discussion and Future Study

From feedback we received after our teaching sessions, we found that students' reception to the subject of algorithm bias ranged from defensiveness and denial of the existence of algorithm bias to openness and acceptance of the existence of such bias. Since the topic ultimately deals with issues of racial, gender-based and other discrimination, a multidisciplinary approach is needed when teaching about algorithm bias. Our assertion is that librarians have a role in partnering with computer science instructors to ensure that students who major in computer science, who will be the primary creators of algorithms as they enter the workforce, can develop an early awareness and understanding of bias in information systems. Further, when the students receive such training, the automated systems they generate will produce more fair outcomes. To this end, algorithm bias instruction could incorporate insights from library science, computer science, medical ethics, and critical theory. The broad aim of algorithm bias instruction should

be to help students recognize and mitigate the systematic marginalization of groups within systems that create information architectures.

We see the module that we developed as an initial step toward building a computer science curriculum where algorithm bias awareness is embedded in coursework throughout the undergraduate experience and continuing into more sophisticated discussions in graduate programs. We saw an opportunity to introduce an instruction module to computer science students about the ethical implications of algorithm bias so that they become aware of biases.

We believe next steps should include an integrative approach that includes library ethics as well as contributions from cultural theorists who have expertise in teaching about race and gender. Because of the capacity of machine learning algorithms to impact large numbers of the world's population, computer scientists have tremendous power to deliver social good but also to cause harm. Our contention is that rather than have students take one or two college courses where algorithm bias is discussed, algorithm bias instruction should be introduced early in a computer science major's study, with continued exposure to the concept in all classes where code with multiple possible outcomes is written.

Moreover, because the workforce that creates computer algorithms that shape our information infrastructure remains predominantly male and white, the skew in the gender and race of programmers affects algorithm design. Wider structural and cultural changes need to occur in the tech industry beyond what is taught in college courses. Open conversations about these changes need to happen in corporate settings alongside a stronger regulatory environment will also help. With regard to legislation, while recent state and federal attempts to pass legislation have failed [18], over a dozen cities have taken the lead only to face barriers. One successful piece of legislation [19] was passed by the city of Portland, Oregon, which banned public and businesses from using facial recognition software, according to The Markup, a non-profit newsroom. Facial recognition software has been critiqued by civil rights activists both for privacy violations and for being error prone in recognizing non-white faces. Also, last December, California Assembly Member Ed Chau [20] introduced AB 13, the Automated Decision Systems Accountability Act of 2021, which aims to end algorithmic bias against groups protected by federal and state anti-discrimination laws. We will continue to watch the public conversations around algorithm bias and software regulation with considerable interest.

Affecting immediate structural changes may be beyond the scope of our instructional modules. Nevertheless, discussions of what practices students need to take into their future jobs are necessary when designing courses on algorithm bias mitigation. To this end, we recommend a multidisciplinary approach involving library and medical ethics when considering semester-long courses for undergraduate computer science majors.

Incorporating Library Science Ethics into Algorithm Bias Instruction

Historically, public libraries, in the context of democratic societies, have advocated for the right of privacy, emphasized intellectual freedom, and promoted community access to knowledge sources. The American Library Association (ALA) is a nonprofit organization based in the United States that promotes libraries internationally. Founded in 1879, the ALA in the United States has a long standing record of taking a stand on issues regarding unequal or restricted access to information caused by segregation, censorship, and state surveillance. In 2003, the ALA [21] passed a resolution opposing the USA PATRIOT Act, calling sections of the law "a present danger to the constitutional rights and privacy rights of library users." The ALA website states the following: “In a political system grounded in an informed citizenry, we are members of a profession explicitly committed to intellectual freedom and the freedom of access to information. We have a special obligation to ensure the free flow of information and ideas to present and future generations.” The ALA [22] also has a Code of Ethics containing eight statements, among them:

> 1. We provide the highest level of service to all library users through appropriate and usefully organized resources; equitable service policies; equitable access; and accurate, unbiased, and courteous responses to all requests.
>
> 8. We do not advance private interests at the expense of library users, colleagues, or our employing institutions.

The strong advocacy for protection from surveillance, both governmental and corporate, as well as the insistence on inclusion and equity in information dissemination makes libraries and librarians strong allies in the quest to mitigate harm caused by algorithm bias. Librarians are also scholars of information theory, in various manifestations, including information seeking behavior. As Kaleev Leetaru [23] points out in his 2019 article “Computer Science Could Learn A Lot From Library And Information Science, “[a]n understanding of the global evolution of how societies have generated, managed, consumed and utilized information throughout history and especially the ways in which societies across the world have differed in their approaches, can offer powerful guidance in the shaping of today’s informational systems.” Our current partnership between a computer scientist and library faculty is then an ideal mix of core computational knowledge and ethical thinking about information, that can lead to a well-designed instructional framework for raising awareness about algorithm bias among college students taking computer science courses.

Another set of concepts we are inspired by is the ACRL Framework for Information Literacy for Higher Education. In 2015, the Association of College and Research Libraries (ACRL) [24] created a framework of six “foundational ideas” to support instruction of information literacy

called the Framework for Information Literacy for Higher Education. According to ACRL, "Information literacy is the set of integrated abilities encompassing the reflective discovery of information, the understanding of how information is produced and valued, and the use of information in creating new knowledge and participating ethically in communities of learning."

One of the six frames, "Information has Value" is a central underpinning of how we would like to target our instruction sessions for computer science students. Increasing students' awareness of the harms of algorithm bias exemplifies how "individuals or groups of individuals may be underrepresented or systematically marginalized within the systems that produce and disseminate information." The systems that produce and disseminate information include for-profit organizations that create biased algorithms. Drawing from this concept of "Information has value," we argue that three dispositions or tendencies will enable computer science students to make informed decisions as creators of future algorithms:

- value the skills, time, and effort needed to produce knowledge.
- see themselves as contributors to the information marketplace rather than only consumers of it.
- examine their own information privilege.

We welcome other librarians and information scientists to join us in our endeavor for dialogue in the academic community about instruction on algorithm bias. We want to note here the work done by Montana State University's Jason Clark [25], associate professor and head of Special Collections and Archival Informatics, Julian Kaptanian, an undergraduate history student, and computer science research assistant, Tyler Bass, on "Unpacking the Algorithms That Shape our User Experience." According to the Montana State University website, "The project includes three main parts, all with a goal of introducing 'algorithmic awareness' as a form of digital literacy: researching algorithms and writing a report for users, developing a teaching tool in order to give transparency to common algorithms, and creating a curriculum and pilot class." We are excited that our colleagues have focused on this issue of significant societal importance and hope to learn from their project and share insights from our work with them when an opportunity presents itself.

In our pilot instructional presentations, we did not specifically incorporate library ethics into our instruction but we think there is scope for more extensive discussion of library and information ethics in future course development. Applying the Embedded EthiCS concept pioneered by Harvard [14], we posit that going forward, there can be a collaborative relationship between the computer science departments and the academic library, where library instructors present to students concepts of information as a public good and fair access to information as a right. Additionally, universities with library science programs, many of them called iSchools, have faculty with significant knowledge about information ethics, such as

Professor Safiya Noble, who teaches in the iSchool at UCLA. We feel that a cross-disciplinary approach to EthiCS should not just focus on the more traditional fields where classical ethics are discussed, such as philosophy, but include knowledge areas such as library and information science. The foundational values of librarianship such as access and the public good [26] could inform principles of inclusive algorithm design, taking it out of the private, proprietary spheres where much of information technology is currently incubated into the public realm.

Incorporating Medical Ethics into Algorithm Bias Instruction

An ethical model drawn from the field of medicine may also be necessary in computer science education as the work computer scientists do continues to impact human beings more widely, with advances in AI. For instance, in the last chapter of *Weapons of Math Destruction*, Cathy O'Neil [3] proposes a Hippocratic Oath for data scientists and writes about how to regulate math models. Although medical students do not literally take pledges and, indeed, the act of taking an oath may have limited effectiveness, medical training focuses assiduously on harm reduction and avoidance and the same focus and intent can be brought into computer science education because even though computer science majors are only working with inanimate and abstract code, the impact of the code is felt by human subjects. Therefore, the idea that people who create algorithms have ethical responsibilities similar to medical practitioners is a concept that could be systematically introduced in computer science education. Possibly, an AI commercial product could go through a review board similar to medical or pharmaceutical research. As O'Neil [3] asserts, "the technology exists! If we develop the will, we can use big data to advance equality and justice."

The ACM/IEEE International [27] Workshop on Software Fairness is compiling a Standard for Algorithmic Bias Considerations. These much needed discussions in professional organizations must also include computer science students as their audience.

Conclusion

As part of our study, we reviewed the background of computer science ethics education and algorithm bias education. We also conducted a student survey to understand their perceptions of algorithm bias, developed an instructional solution, and, finally, proposed pathways for additional research for an interdisciplinary solution to teaching about algorithm bias. Our multidisciplinary, holistic instructional methods prepare students to survey the landscape, embrace ambiguity, recognize root causes, and prioritize reducing algorithm bias and widespread harms. We hope that our ongoing project will serve as an initial resource for a comprehensive computer ethics education regarding algorithm bias. We are also encouraged that both instructors and students are becoming actively involved in finding solutions to the problem of how to avoid hardwiring societal bias into our computing machines. As Ashley

Shadowen, a student at CUNY sums up in her Masters' thesis, "Machine ethics is a complicated and multifaceted problem. But if we get it right, we will unleash the full benefit of machine learning for humankind." [28]

References

[1] Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM Transactions on Information Systems (TOIS), 14(3), 330–347. https://doi.org/10.1145/230538.230561

[2] S. Noble, *Algorithms of oppression: How search engines reinforce racism.* New York: NYU Press, 2018.

[3] C. O'Neil, *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown 2016.

[4] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, vol. 366, no. 6464, pp. 447–453, 2019, https.//doi.org/10.1126/science.aax2342.

[5] J. Angwin, J. Larson, S. Mattu, & L. Kirchner, "Machine Bias," *ProPublica*. [Online]. Available: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing . [Accessed: 9-Mar-2021]

[6] S. Athey, "A comparison of experts' and high tech students' ethical beliefs in computer-related situations," *J Bus Ethics*, vol. 12, no. 5, pp. 359–370, May 1993, doi: 10.1007/BF00882026.

[7] S. Bowers, E. Maccarone, and G. Ricco, "On the Integration of Ethical, Legal, and Societal Issues into a Computer Science Senior Design Capstone Program," in *2016 ASEE Annual Conference & Exposition Proceedings*, New Orleans, Louisiana, 2016, p. 25826, doi: 10.18260/p.25826 [Online]. Available: http://peer.asee.org/25826. [Accessed: 9-Mar-2021]

[8] I. Hilliger, A. Strello, F. Castro, and M. Pérez-Sanagustín, "Are All Engineering Students Capable of Recognizing Ethical and Professional Issues? An Assessment Approach to Engineering Ethics," in *2017 ASEE Annual Conference & Exposition Proceedings*, Columbus, Ohio, 2017, p. 27610, doi: 10.18260/1-2--27610 [Online]. Available: http://peer.asee.org/27610. [Accessed: 9-Mar-2021]

[9] A. Hedayati Mehdiabadi, "The Ethical Judgement Processes of Students in Computing: Implications for Professional Development," in *2018 ASEE Annual Conference & Exposition Proceedings*, Salt Lake City, Utah, 2018, p. 31099, doi: 10.18260/1-2--31099 [Online]. Available: http://peer.asee.org/31099. [Accessed: 9-Mar-2021]

[10] K. Fu, R. Kirkman, and B. Lee, "Teaching Ethics as Design," *2017 Fall AEE Journal*, vol. 6, no. 2, p. 31326, 2017, doi: 10.18260/3-1-170.370.620-31326.

[11] T. DiStefano *et al.*, "Ethics For First Year Engineers: The Struggle To Build A Solid Foundation," in *2005 Annual Conference Proceedings*, Portland, Oregon, 2005, p.

10.589.1-10.589.9, doi: 10.18260/1-2--15449 [Online]. Available: http://peer.asee.org/15449. [Accessed: 9-Mar-2021]
[12] M. J. Quinn, *Ethics for the information age*, 7th edition. Boston: Pearson, 2016.
[13] T. Bynum, "Computer and Information Ethics," Aug. 2001 [Online]. Available: https://plato.stanford.edu/archives/sum2020/entries/ethics-computer/. [Accessed: 9-Mar-2021]
[14] B. J. Grosz *et al.*, "Embedded EthiCS: integrating ethics across CS education," *Commun. ACM*, vol. 62, no. 8, pp. 54–61, Jul. 2019, doi: 10.1145/3330794.
[15] T. Abate, "How the Computer Science Department is teaching ethics to its students," *Stanford School of Engineering*, 20-Aug-2020. [Online]. Available: https://engineering.stanford.edu/news/how-computer-science-department-teaching-ethics its-students. [Accessed: 9-Mar-2021]
[16] S. Fu, S. M. Cutchin, K. Howell, and S. Ramachandran, "Full Paper: Algorithm Bias: Computer Science Student Perceptions Survey," presented at the Proceedings of the 2020 ASEE PSW Section Conference, canceled, 2020 [Online]. Available: https://peer.asee.org/full-paper-algorithm-bias-computer-science-student-perceptions-survey. [Accessed: 28-May-2021]
[17] A. Smith,"Public Attitudes Toward Computer Algorithms," *Pew Research Center: Internet, Science & Tech*, 16-Nov-2018. [Online]. Available: https://www.pewresearch.org/internet/2018/11/16/public-attitudes-toward-computer-algo r ithms/. [Accessed: 9-Mar-2021]
[18] "Legislation Related to Artificial Intelligence." [Online]. Available: https://www.ncsl.org/research/telecommunications-and-information-technology/2020-le g islation-related-to-artificial-intelligence.aspx. [Accessed: 9-Mar-2021]
[19] "Police Say They Can Use Facial Recognition, Despite Bans – The Markup." [Online]. Available: https://themarkup.org/news/2021/01/28/police-say-they-can-use-facial-recognition-despi t e-bans. [Accessed: 9-Mar-2021]
[20] "Legislator Detail | BillTrack50." [Online]. Available: https://www.billtrack50.com/legislatordetail/17180. [Accessed: 9-Mar-2021]
[21] K. de la P. McCook, *Introduction to public librarianship*, 2nd ed. New York, NY: Neal-Schuman Publishers, 2011, pp. 63–64.
[22] American Library Association. "Professional Ethics," *Tools, Publications & Resources*, 19-May-2017. [Online]. Available: http://www.ala.org/tools/ethics. [Accessed: 9-Mar-2021]
[23] K. Leetaru, "Computer Science Could Learn A Lot From Library And Information Science," *Forbes*. [Online]. Available: https://www.forbes.com/sites/kalevleetaru/2019/08/05/computer-science-could-learn-a-l ot-from-library-and-information-science/. [Accessed: 9-Mar-2021]
[24] Association of College and Research Libraries. "Framework for Information Literacy

for Higher Education," 09-Feb-2015. [Online]. Available: http://www.ala.org/acrl/standards/ilframework. [Accessed: 9-Mar-2021]
[25] J. A. Clark, *jasonclark/algorithmic-awareness*. 2021 [Online]. Available: https://github.com/jasonclark/algorithmic-awareness. [Accessed: 9-Mar-2021]
[26] American Library Association Council, "Core Values of Librarianship," Advocacy, Legislation & Issues, 26-Jul-2006. [Online]. Available: http://www.ala.org/advocacy/intfreedom/corevalues. [Accessed: 19-Apr-2021]
[27] A. Koene, L. Dowthwaite, and S. Seth, "IEEE P7003™ standard for algorithmic bias considerations: work in progress paper," in *Proceedings of the International Workshop on Software Fairness*, Gothenburg Sweden, 2018, pp. 38–41, doi: 10.1145/3194770.3194773 [Online]. Available: https://dl.acm.org/doi/10.1145/3194770.3194773. [Accessed: 9-Mar-2021]
[28] A. Shadowen, "Ethics and Bias in Machine Learning: A Technical Study of What Makes Us 'Good,'" *Student Theses*, Dec. 2017 [Online]. Available: https://academicworks.cuny.edu/jj_etds/44. [Accessed: 19-Apr-2021]

Appendix A: Survey Questions

1. How often do you use computer search engines?
2. Which search engine do you use most often?
3. Which search engine do you use most often?
4. Please rate your level of satisfaction with your search results (i.e. how you find, review, and use your results).
5. Your browser history reflects your personal identity.
6. Search engine results are good enough for every day questions.
7. Search engine results are accurate.
8. Search engine results are complete.
9. Search engine results are trustworthy.
10. I have used the auto-fill feature in Google.
11. I have unexpectedly encountered racially offensive search results when doing a routine search.
12. I have unexpectedly seen results that are derogatory to people with disabilities when doing a routine search.
13. Computer science students are well trained about the ethical impact of their technology design choices for computer algorithms.
14. Private companies need to self-regulate AI algorithms for improving society.
15. Technology will reduce discrimination in society.
16. The services AI scientists build are free of discrimination.

17. Search engines display objective results.
18. AI algorithms will improve over time.
19. AI and machine learning algorithms are bias free.
20. Engineers can create biased algorithms.
21. Government needs to regulate search engine companies to ensure fair search results.

Appendix B: Survey Summary

1. Google is the dominant search engine (97%) used by the respondents with over 88% of them using computer search engines seven days a week.
2. An overwhelming agreement (92%) exists that AI algorithms will improve over time with over 76% “Agree” and “Strongly agree.”
3. On computing the difference of opinion between “AI and machine learning algorithms are bias free” and “Engineers can create biased algorithms”, there was a slight positive tilt that people who agree that algorithms are bias free are also agreeing that engineers can create biased algorithms, with the difference being less pronounced in PhDs and more in master’s students. On average, people are willing to acknowledge bias is present when it is clear that people are involved in the creation of the algorithm.
4. Graduate students were more in positive agreement for “Government needs to regulate search engine companies to ensure fair search results" while undergraduates had neutral opinions. There is clear positive agreement across all levels for “Private companies need to self-regulate AI algorithms for improving society.”
5. Boise State University students were neutral on “Technology will reduce discrimination in society.” while University of Southern California students are in slight agreement. On government regulation, Boise State University students are in slight disagreement while California State University, Los Angeles and University of Southern California students are in slight agreement.
6. For “I have unexpectedly seen search results that reflect gender stereotypes when doing a routine search,” 42.9% of females report as true compared to 34.4% of males.
7. For “I have unexpectedly seen search results that reflect gender stereotypes when doing a routine search,” Latinx reported the highest at 28.8% with the least being those who identified as Other at 16.1%.
8. Our survey did not directly ask students whether they had heard of or were aware of the concept of algorithm. bias. The survey questions were to determine if students had perceived any bias in their daily interaction with search engines and AI.
9. When it came to search engines, most students did not report overt bias but negative reporting was higher among women and Latinx students. The count of black students in the survey was too low for us to make a demographic inference about their experience of bias.