Final Group Report - HUM433

Sara Djambazovska; Yutong Peng; Chenkai Wang
EPFL, May. 2021

An implicit bias is an unconscious association, belief, or attitude toward any social group. Due to implicit biases, people may often attribute certain qualities or characteristics to all members of a particular group, a phenomenon known as stereotyping [5].

What exactly you want students to learn and how exactly do you want them to learn it (with an emphasis on what has changed since your interim report and why)?

We chose to incorporate Design Ethics, in particular, **to reduce implicit bias among first-year Computer Science (CS) students**, by building empathy and concerns for social justice. Students can learn to think not only about *what* technology they could create, but also *whether* they should create it and *who* will be the end-user.

We are proposing an interactive way of teaching, incorporated into the weekly exercise sessions. The students will be given different **ethical algorithms to implement**, which besides testing and improving their technical skills should get them to reconsider their moral values and biases. The artifact attached shows an example of such a simple game - algorithm for implementing the decision for self-driving cars (coded in Python, see artifact or use this Colab link). In this particular case, the algorithm must choose to save/kill a group of people - the passengers or pedestrians, so the students need to add/code a set of rules (based on the physical/social characteristics of the people) deciding which group is run over. Later there will be a discussion on how they felt implementing this game.

This type of activity will be performed repeatedly, at least twice per semester, as suggested by Hess and al. [6], spaced and repeated practice is a good approach to develop empathic perspective-taking. The students will be assessed based on their ability to design and code a program, and the ethical aspects can be assessed and

used as a bonus. The assessment of ethical aspects can be done by presenting the students with a case study at the beginning and end of the semester, for which they would need to write a short reflection, grading their ethical reasoning development.

Describe the prototyping studies you completed (at least one in detail), what questions were they intended to answer and what information did you get from them?

In order to know whether it is useful to address the implicit bias of bachelor computer sciences students, and whether the way we are proposing to address it will arouse some interest among the students, we sent out a questionnaire. We received an overwhelming response, with more than 300 students answering the questionnaire (see Appendix).

The first questions we put in our questionnaire are the basic information about the student and if the students are aware of what is implicit bias. This can already tell if it is useful that we develop a tool to address implicit bias.

As responses, about half of participants said that they do not know what implicit bias is. There was also an even distribution of participants from Bachelor semester 2, 4, and 6 and master students from semester 2 and 4. It is interesting to notice that even some MA4 students do not know what implicit bias is. Hence these two facts confirm the need to make students aware, then reduce their own implicit bias.

Then, we gave a definition of implicit bias so that the participants can keep it in mind while answering the remaining questions.

We included three “yes or no” type questions, where we were interested to know if the students have ever realized that a decision they made was biased. We also asked about whether the students are willing to address their own bias, probably by having an integrated ethics module. Although the majority of students (around 90%) answered “yes” to the first two questions, namely realizing that their decisions could have been biased and addressing their own bias, only 30% said “yes” to an integrated ethics module. 40% went for “maybe”.
Through these answers, it can be seen that the students are willing to learn how to

reduce their own bias, but not under the form of a course completely devoted to this aim. Therefore, our proposal to include the teaching of reducing bias in several exercise sessions repeatedly throughout the semester might be accepted by the students, which is confirmed in one question we asked afterward.

In a third part, we put 12 “Strongly disagree - Strongly agree” type questions, where we inquired about whether participants think that implicit bias can be “corrected” by training; whether the abstraction computer sciences offer while creating new algorithms leaves the students unaware, and thus not feeling responsible for the possible consequences of their work on the end user. There was also a question on if students are aware of their own biases and another on if they think that these biases would be translated into the algorithm created. There were questions about if the participants think there is a gender bias and whether diversity (racial & social) is important in the field of technology. Moreover, we wanted to know if the students think that eliminating bias is as important as teaching how to code algorithms, and if they are willing to participate in in-class debate (as we proposed initially as the method to address implicit bias).
Considering the responses, most people went for “agree” or “strongly agree” on the questions concerning awareness, correction of, translation into algorithms, and the importance of addressing implicit bias. They also agree on the importance of diversity in the field of technology and gender bias. But about 66% still think that eliminating bias is not as important as teaching coding algorithms. Thus, it enforces the fact our tool would be useful to implement.
However, as the majority was not very prone to participate in in-class debate, we switched our method of addressing bias to the implementation of ethical algorithms.

Apart from the responses, just looking at the basic information about the participants, there is a clear gender bias: 77% were male students. Probably we should first start by addressing the bias that females are less likely to succeed in STEM branches.

Do you think your design will achieve the above learning goals (evaluating evidence for and against and describing how your thinking has developed since late March)?

Arguments for:

Given the summary of responses we obtained from EPFL Computer Science students, we are positive that the target learning goals (reducing the implicit bias among CS students) can be obtained by the interactive game developing approach proposed.

This is based on the premise that implicit bias is like a habit that can be reduced through “a combination of **awareness** of implicit bias, **concern about the effects of that bias**, and the application of **strategies to reduce bias**.” [1].

Research also finds that the approach proposed, becoming aware of your own implicit biases and then receiving advice and training in managing them, can make a significant impact. A gender bias study [4] found that when at least a quarter of the faculty attended only one interactive workshop on the topic of gender bias, there were significant increases in self-reported actions to promote gender equity at three months. Having the whole class participate by working on the proposed algorithmic ethics game is expected to bring satisfying results in terms of raising awareness of implicit bias, and further leading to its reduction. Additionally, the suggested Interactive way of teaching is evaluated as one of the best approaches for teaching ethical issues in CS in [7].

From the questionnaire, we obtained the surprising information that about half of the CS students at EPFL do not even know what implicit bias means. So, a big leap forward would be to just get them **aware** and thinking about the concept and effect of the implicit bias, by showing them the results of their (biased) decisions once the game algorithm is randomly run. Furthermore, about 90% of the students declared that they are willing to address their own biases, which is our main goal. Again, when asked if engineers *need* to become aware of their own implicit biases the vast majority agreed (only 5% disagreed).

As tested in [2] we believe that interactive training (through workshops, talks, ethics exercises…) can reduce the implicit bias and about 57% of CS students (strongly) agreed with this. The majority of the students **showed concern about the effects of the engineer’s implicit bias**, (70% of students) by stating that reducing this bias among developers will benefit the users, giving them a more equal experience, and

more than 82% agree that the engineer's implicit bias might be translated into the algorithm they will create.

More than half of the students agreed that the abstraction that CS offers while creating algorithms leaves the student developers unaware of the possible effect their implementation will have on the end-users, never directly feeling responsible for the possible consequences of their work. Upon implementing the decision algorithm in the prototype game that we propose, the students get immediate feedback on how their (unconsciously) biased decisions will affect the end-users.

Arguments against:

We might encounter the problem of lack of compassion, as students, unaware of their own implicit bias or the importance of overcoming it, might focus only on their technical skills - getting the algorithm to work and not caring about the end result.

A lot of the students do not see the importance of eliminating implicit bias among developers, as only 33% believe that it is as important as teaching CS algorithms, and only 30% strongly feel the need for an integrated Ethics CS course in their curriculum.

We are also unable to assure that the awareness and knowledge gained by co-developing this game will remain with these future software engineers when they will be building new products. As this meta-analysis from about 500 studies on implicit measures shows [3], the effects of the implicit bias change are often relatively weak and short-term. Furthermore, they found that "changes in implicit measures are possible, but those changes do not necessarily translate into changes in explicit measures or behaviour.".

Do you think your tool will/would actually be used in practice (evaluating evidence for and against and describing how your thinking has developed since late March)?

Arguments for:

Based on the responses we obtained from EPFL CS students, we positively reckon that our method with a few adjustments can be generally used in practice. Additionally, a game developing approach will be adopted instead of the initially proposed debate. Games, such as the artifact one, will be designed to implement different algorithms related to ethics problems.

Why would our proposal actually be used in practice? We would divide our demonstration into three parts, which are
1) interviewees admit to the existence of prejudice;
2) interviewees identify with the benefits of mitigating implicit biases;
3) interviewees would like to correct their implicit biases and make a change.
These reasons are connected and demonstrate our views logically. In the following part, the statistics from our responses summary will support our perspective.

First of all, when asked "Have you ever made a decision that you later realized was biased?", 87.9% of interviewees answered “Yes”. This indicates nearly ninety percent of the people admit their implicit biases, though after things occur. Surprisingly, 65.7% of interviewees (35.3%" strongly agree" and 30.4% "agree") recognize that there is a gender bias in the field of Technology (developers). This illustrates the students’ awareness of the existence of (implicit) biases.

Secondly, half of the interviewees(50.7%) strongly agree that engineers need to become aware of their own implicit biases, while 34.3% agree. And 56% of interviewees (13.7%" strongly agree" and 43.8% "agree") confirm that implicit biases can be “corrected” by training. Furthermore, 70% of them are aware that reducing the implicit bias among developers can benefit the users, giving them a better (equal) user experience. This is the crucial step that realizing the benefits to eliminate implicit biases can prompt students to participate in our project.

Furthermore, we are excited that the questionnaire indicates that 89.5% of interviewees are willing to address their implicit biases. This confirms that the majority of students from EPFL CS might be interested in engaging in this interactive game development. What's more, half of the interviewees (24.8% "strongly agree" and 26.1% "agree") are positive that EPFL should integrate ethics teaching or create a separate module teaching ethics in CS.

Lastly, our design can not only enable students to solve technical-level problems but also get them to consider ethical aspects. In general, using our designed tool (implementing different ethical algorithms) would enhance the students' coding skills. So even the students who are not interested in the ethics behind it would still be driven to attempt and develop an algorithm as it would be an exercise using the technical material learned in class.

Arguments against:

Our initial proposal was to hold an in-class debate. Thanks to the prototyping study, we realized that most students might not be prepared to debate. The results highlight that only 13.7% of students are strongly agreeing to attend an in-class debate, and 25.8% of them agree to do that. At the same time, the majority of interviewees are centralized at "Strongly disagree" and "disagree". The above statistics indicate that the debate method might not work as we predicted in practice. There could be a lot of reasons leading to this consequence: 1) EPFL is an international school and the class is taught in French, debating in French might discourage non-francophone students from attempting. 2) Students might not feel interested in participating in the debating parts.

Another issue could be that the professor, M. Schinz, with a CS background, might not be competent enough to grade their reflections work and/or lead the debates. Additionally, while students may reflect on the issue proposed during the lecture, however, outside the academic framework, in everyday life, they are not considering the ethical issues. It is rather hard to develop a habit, therefore the professor might consider it as some kind of a waste of time.

Recommendations for future development of the tool

In order to introduce all students to the basic debating techniques, live debate presentations can be done, perhaps by inviting students from MUN-Model United

Nations EPFL (inviting a guest lecturer from another field was considered an effective solution in [7]).

Making the worth of the discussion as a bonus to the grade: students need to participate and write the reflection to get the bonus (the reflections do not need to be very neatly stated).

Co-designing the class with the help of a professional from the field of philosophy, as suggested at Harvard [8].

Maybe only one semester is not enough for reducing one's implicit bias, hence, one should seek cooperation among different professors across different semesters.

In order to make the students realize that ethics is a real issue of daily life, especially when they are designing new things, the gaming subjects should be based on real examples.

Conclusion

The numerous responses we received from the students show their eagerness to become aware and reduce their own biases, as well as their concern for the effects of this bias. It allows us to make the assumption that our proposed strategy (an ethics game development instead of the initial idea of weekly debate) will be used in practice achieving the expected results, namely making students aware and thus reducing their own bias.

Reference

[1] Devine, Patricia G et al. “Long-term reduction in implicit race bias: A prejudice habit-breaking intervention.” *Journal of experimental social psychology* vol. 48,6 (2012): 1267-1278. doi:10.1016/j.jesp.2012.06.003

[2] Stone, J., Moskowitz, G. B., Zestcott, C. A., & Wolsiefer, K. J. (2020). Testing active learning workshops for reducing implicit stereotyping of Hispanics by majority and minority group medical students. *Stigma and Health, 5*(1), 94–103. https://doi.org/10.1037/sah0000179

[3] Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology, 117*(3), 522–559. https://doi.org/10.1037/pspa0000160

[4] Carnes M, Devine PG, Baier Manwell L, Byars-Winston A, Fine E, Ford CE, Forscher P, Isaac C, Kaatz A, Magua W, Palta M, Sheridan J. The effect of an intervention to break the gender bias habit for faculty at one institution: a cluster randomized, controlled trial. Acad Med. 2015 Feb;90(2):221-30. doi: 10.1097/ACM.0000000000000552. PMID: 25374039; PMCID: PMC4310758.

[5] A, J. T. J. , B, L. A. R. , C, I. V. B. , D, D. R. C. , E, N. D. , & F, J. G. , et al. (2009). The existence of implicit bias is beyond reasonable doubt: a refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. Research in Organizational Behavior, 29, 39-69.

[6] Hess, JL, Beever, J, Zoltowski, CB, Kisselburgh, L, Brightman, AO. Enhancing engineering students' ethical reasoning: Situating reflexive principlism within the SIRA framework. J Eng Educ. 2019; 108: 82– 102. https://doi.org/10.1002/jee.20249

[7] Schulze, Kay G. and Frances Grodzinsky. "Teaching Ethical Issues in Computer Science: What Worked and What Didn't." ACM SIGCSE Bulletin 28.1 (1996): 98-101.

[8] Barbara J. Grosz, David Gray Grant, Kate Vredenburgh, Jeff Behrends, Lily Hu, Alison Simmons, and Jim Waldo. 2019. Embedded EthiCS: integrating ethics across CS education. Commun. ACM 62, 8 (August 2019), 54–61. DOI:https://doi.org/10.1145/3330794

Appendix

EthiCS

Questions Responses 306

306 responses

Not accepting responses

Message for respondents

This form is no longer accepting responses

Summary Question Individual

Name (Optional)

75 responses

Dubravka Kutlesic

Nevena Dresevic

Mikhail

Bergeron

Richard

Lilia

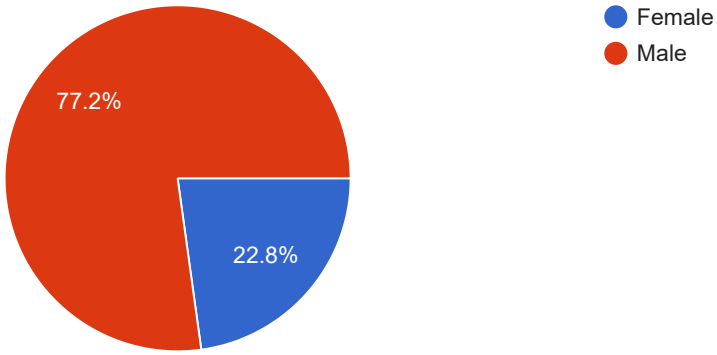
Alain Mérillat

Mark

Stephane

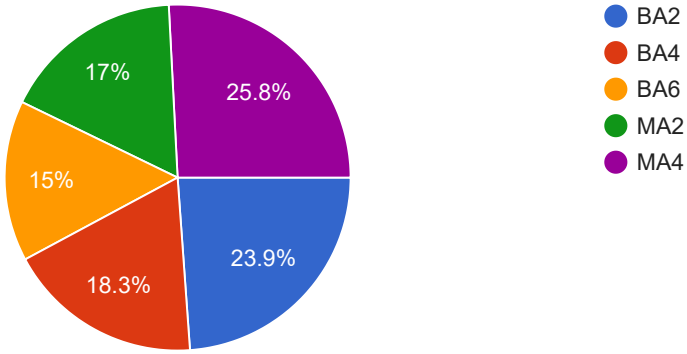
Gender

263 responses



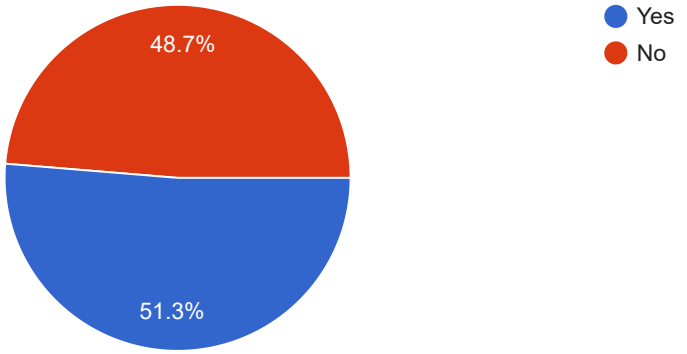
Semester currently enrolled in

306 responses



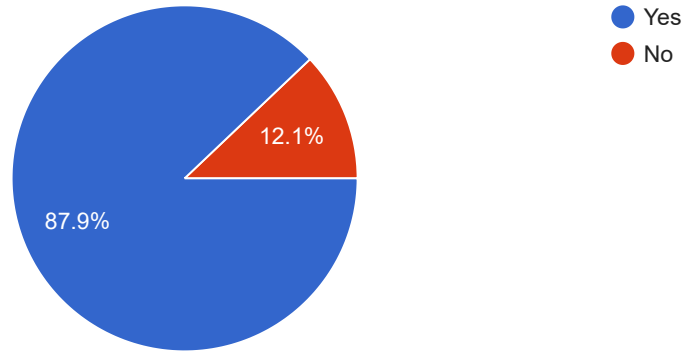
Do you know what implicit bias means?

306 responses



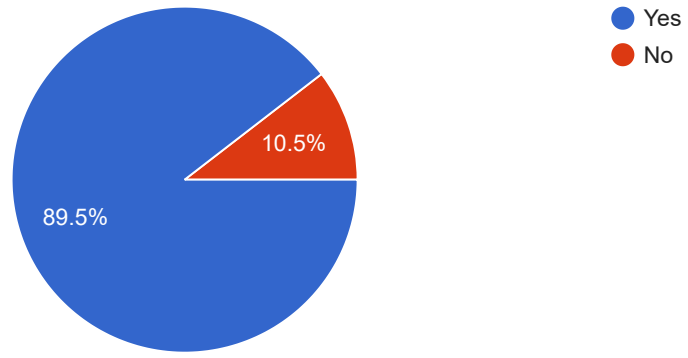
Have you ever made a decision that you later realized was biased?

306 responses



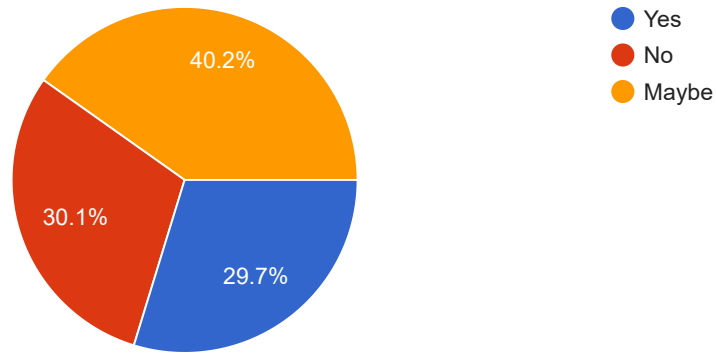
Are you willing to address your own (implicit) biases?

306 responses



Do you feel the need for an integrated Ethics CS module in your curriculum?

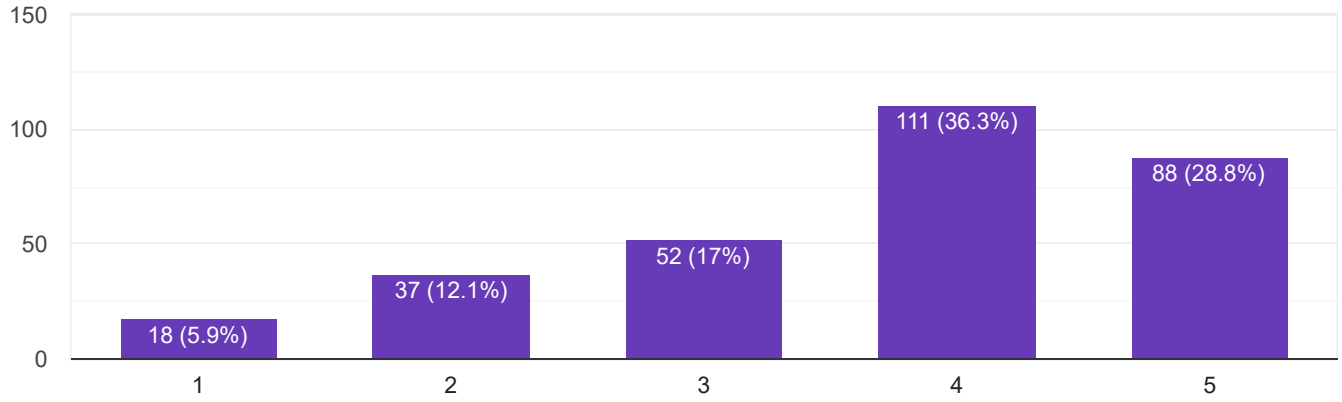
306 responses



Untitled Section

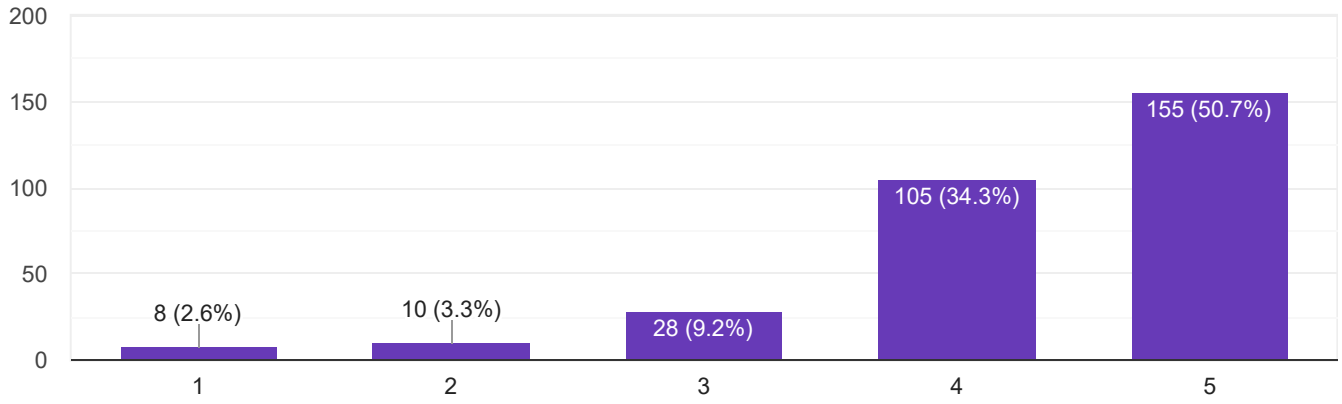
The engineer's implicit bias might be translated into the algorithm they create.

306 responses



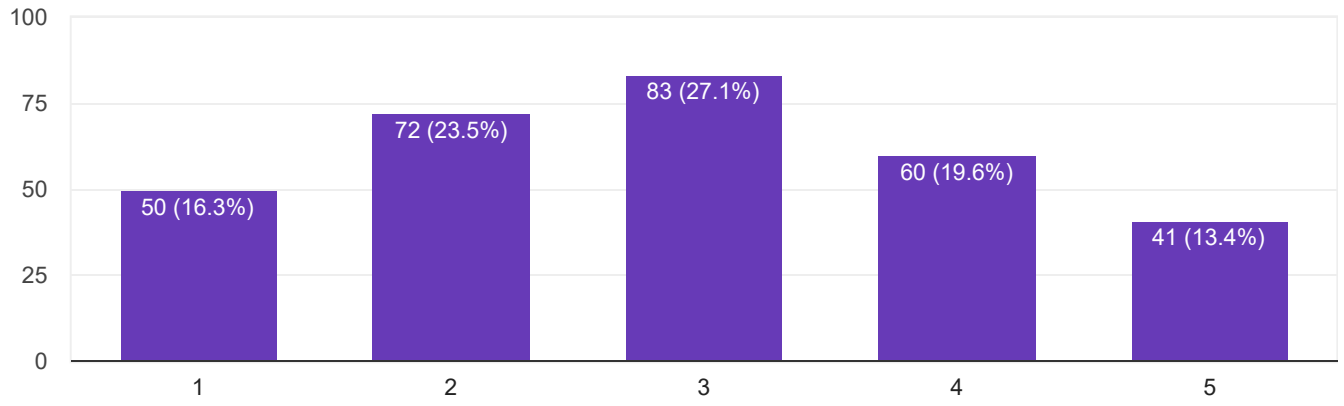
Engineers need to become aware of their own implicit biases.

306 responses



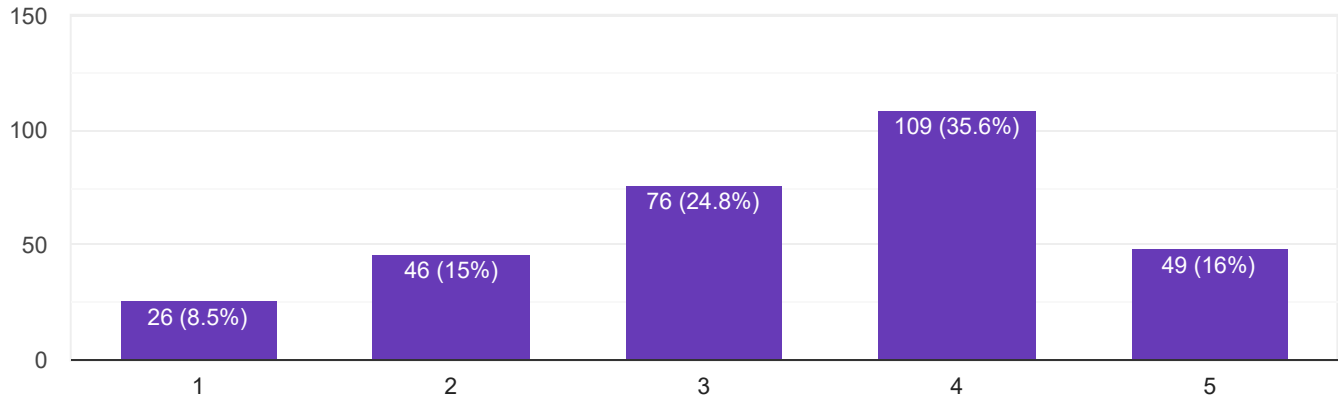
Eliminating implicit bias in coding is as important as teaching CS algorithms.

306 responses



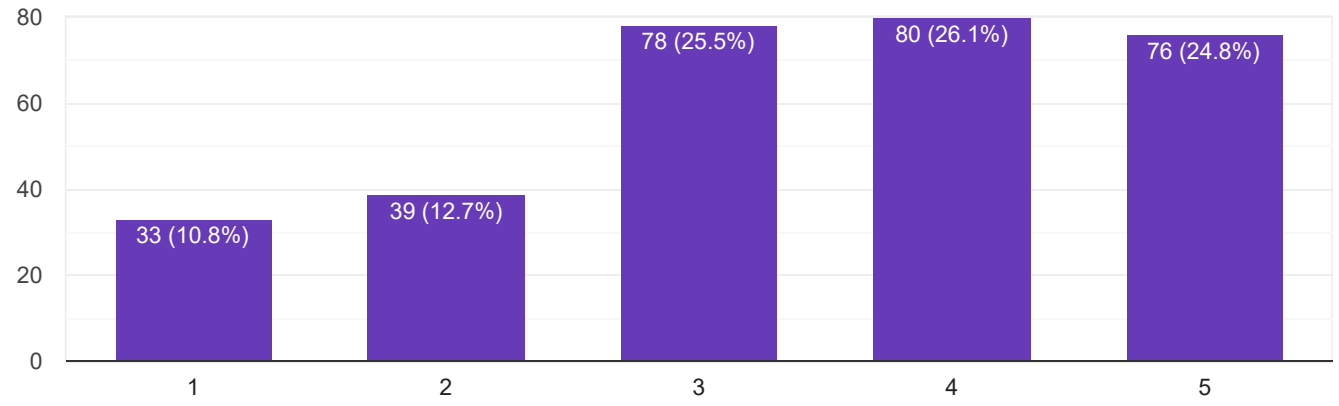
The abstraction that CS offers while creating algorithms leaves the student developers unaware of the possible effect their implementation will have or could have on the end-users, never directly feeling responsible for the possible consequences of their work.

306 responses



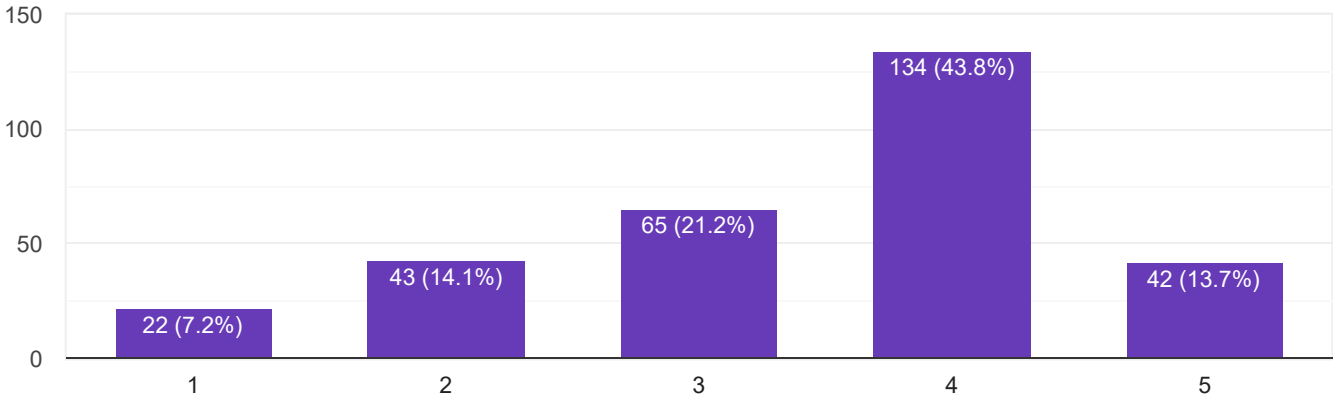
EPFL should integrate ethics teaching or create a separate module teaching ethics in CS.

306 responses



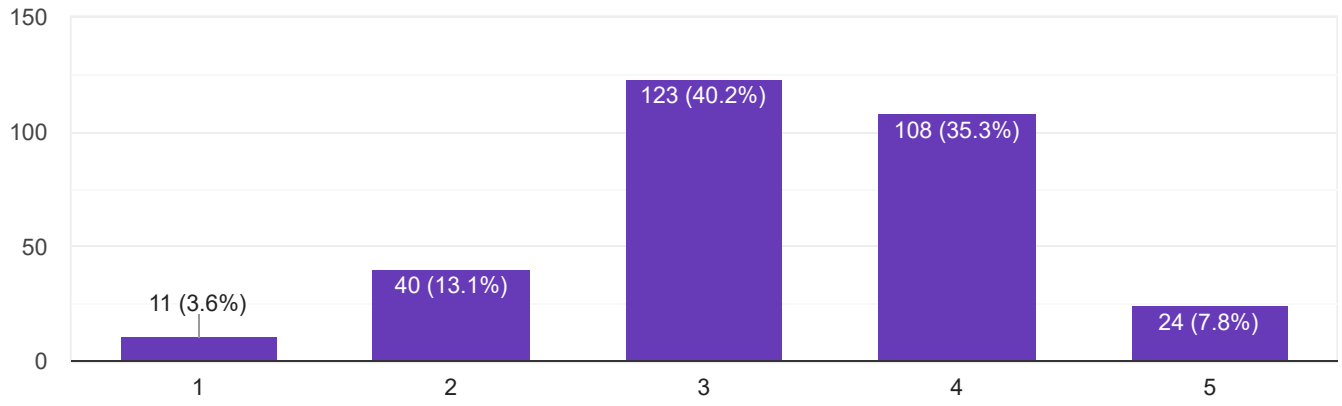
Implicit biases can be “corrected” by training.

306 responses



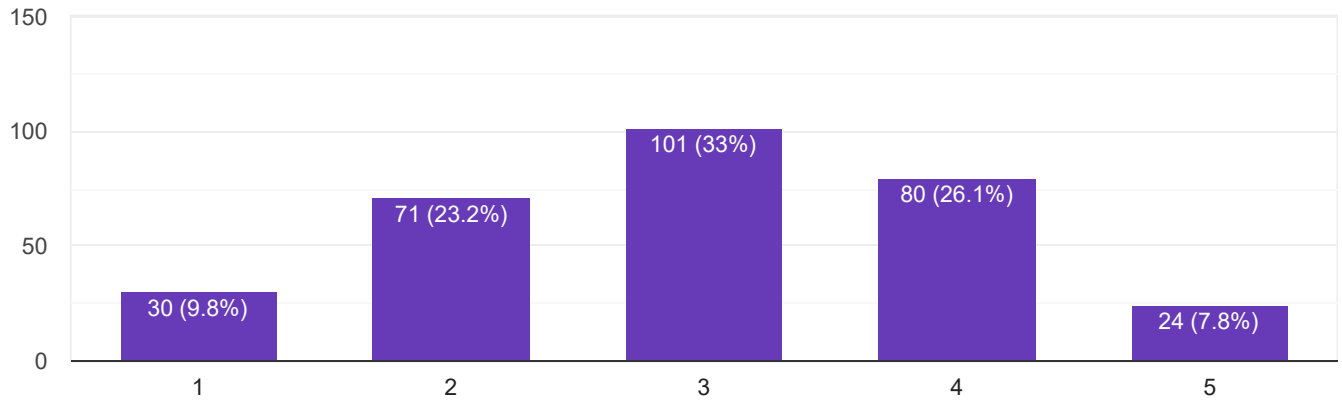
I am aware of my own biases.

306 responses



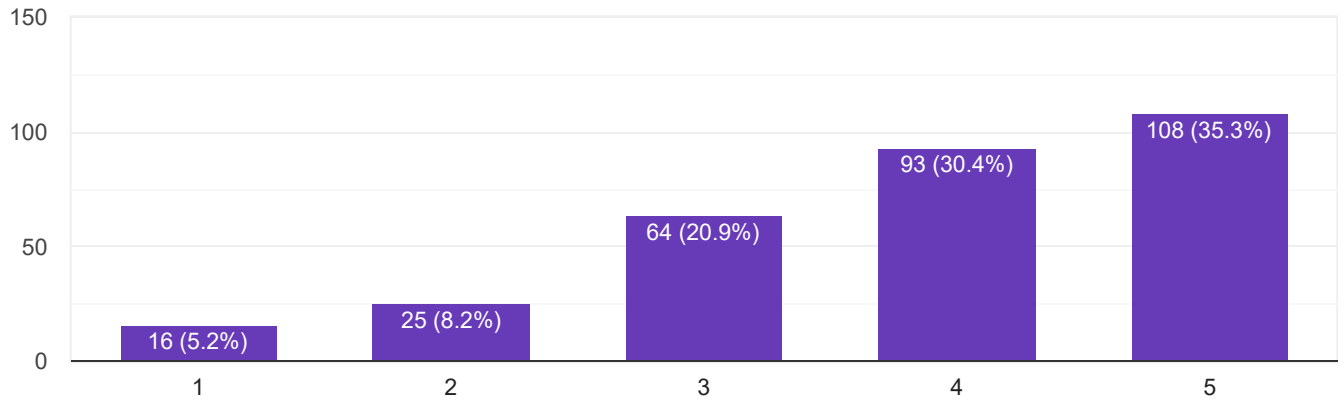
Whenever I am designing a new algorithm I consider the whole range of end users and what effect/impact (positive or negative) it could have on them.

306 responses



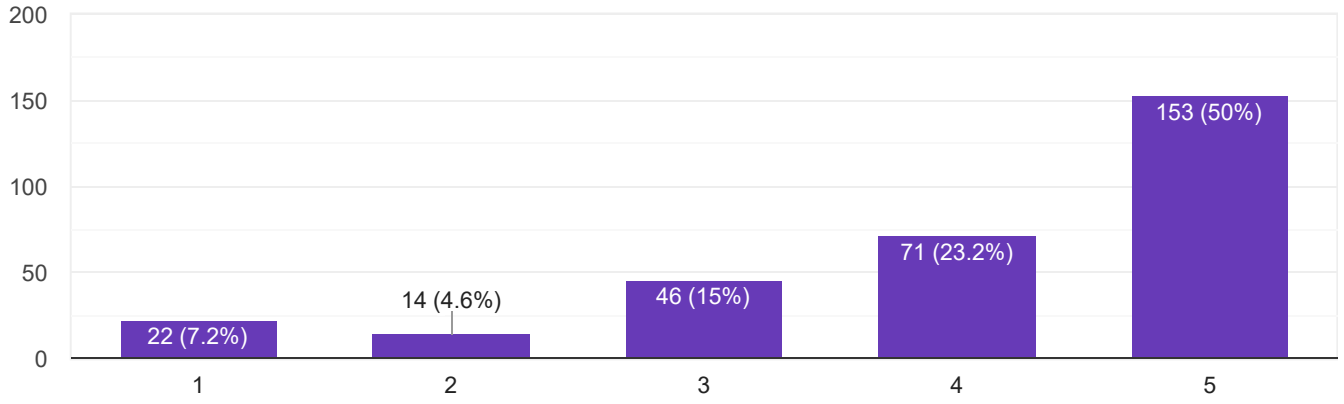
There is a gender bias in the field of Technology (developers).

306 responses



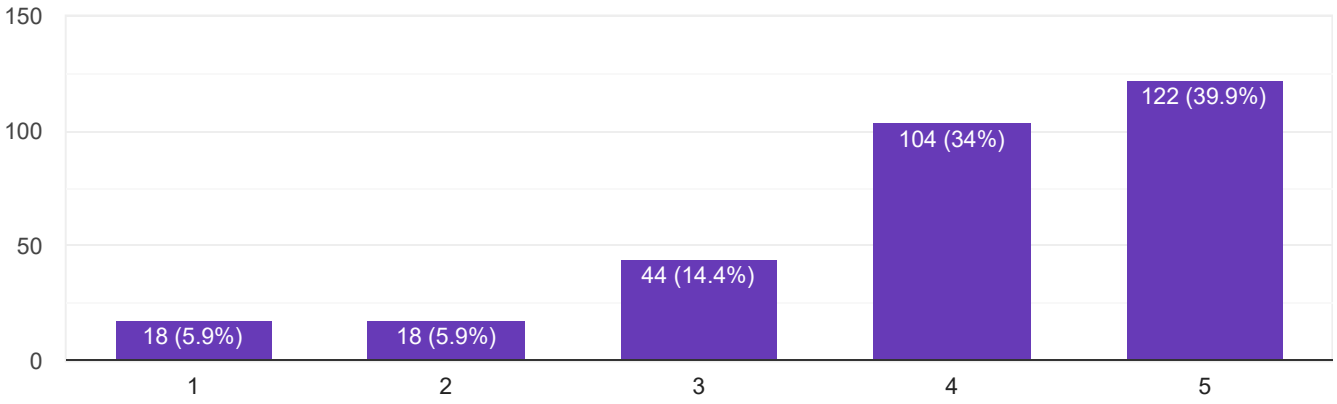
Racial and social diversity is important in the field of Technology (developers).

306 responses



Reducing the implicit bias among developers can benefit the users, giving them a better (equal) user experience.

306 responses



How likely are you to attend an in-class debate (upon receiving proper training if needed)?

306 responses