# Training Socially Conscious Engineers: How STS Application Minimizes Discriminatory Technologies

**Bryn Seabrook**
*University of Virginia*

## Abstract

In the United States, we are constantly searching for new forms of innovation, particularly in relation to technology. However, not all new technologies are inclusive. Motion sensor sink faucets that cannot detect pigment in a user's hand, or Snapchat filters that cannot detect a user's face—these are a small sample of technologies that are discriminatory. The current engineering design process does not account for the need to identify the possible biases or flaws that could contribute to discrimination. In an effort to create technologies that are more inclusive of the entire population, this research paper investigates the following question: How can the field of science and technology studies (STS) be applied to the engineering design process to minimize technical design flaws that ultimately lead to discriminatory technologies? This research uses different STS theories to analyze the engineering design process, ultimately suggesting how the field of STS can contribute positively to the overall design process.

## Keywords

Engineering design
Discriminatory technologies
Inclusive technology
Science, technology, and society
Machine learning

## A Racist Soap Dispenser

At a Marriott hotel in Atlanta, Georgia, guests of the Dragon Con sci-fi and fantasy convention quickly discovered an issue with an automatic soap dispenser. An African American guest was not able to activate the sensor to detect his hand, while a white guest was able to use the dispenser without any delay. British company Technical Concepts designed this soap dispenser using an infrared detection system. Users with lighter skin tones reflect an infrared signal back into the sensor, signaling the dispenser to release soap. However, those users with darker skin tones absorb the infrared signal instead of reflecting it, effectively eliminating these users' ability to get soap from the dispenser.[1] This flaw in the design of the soap dispenser demonstrates the importance of engineers using a holistic approach to consider different kinds of users. This research paper investigates the following question: How can the field of science and technology studies (STS) be applied to the engineering design process to minimize technical design flaws that ultimately lead to discriminatory technologies? The analysis of available data suggests that there should be an additional step in the current design process that specifically addresses inclusivity, thus training engineers to be more socially conscious. This paper uses different STS theories to analyze the engineering design process, ultimately suggesting how the field of STS can contribute positively to the overall design process.

## Placing Engineering Design in Context

One STS concept that helps answer this research question is the Social Construction of Technology (SCOT) framework. The SCOT framework encapsulates the sociotechnical processes that drive technological change by considering system stakeholders and influences as interacting entities.[2] Originally introduced by British sociologist, Trevor Pinch, and Dutch philosopher, Wiebe Bijker, SCOT outlines four main components to organize this analysis. These components are as follows: interpretive flexibility, relevant social group, closure and stabilization, and wider context. These components will help highlight the benefits and risks of the engineering design process, and highlight recommendations to mitigate issues of technically flawed technologies.

This paper analyzes the research question by using Rittel and Webber's wicked problem framing to account for the complex network of social, economic, and political factors associated with technology. The wicked problem framework outlines an approach for inherently unsolvable challenges with complex systems that prevent the analyst from defining strict problem boundaries.[3] Rittel and Webber (1973)[3] also describe the interrelatedness of wicked problems as an important contributor to their complexity; attempting to solve revitalization would subsequently affect other sociopolitical wicked problems. Some scholars have claimed that the wicked problem interpretation that "the planner has no right to be wrong" no longer holds today, given that rapid prototyping and "fast failure" are touted as the keys to good design strategy.[4] Even so, the core facets of the framework regarding solution quantification and stopping criteria are extremely applicable to this analysis. The wicked problem framework helps meter the sweeping implication of the research question by acknowledging factors that can impede the realization of theoretical projections. The wicked problem lens also provides a means to compare the relative power of the stakeholders in this system, one commonly cited drawback of the social construction of technology.

## Examples of Discriminatory Technologies

A subset of artificial intelligence (AI), the field of machine learning is creating the ability for a computer to understand the real world. Machine learning enables algorithms to learn from patterns in data to make reliable decisions.[5] The first example of machine learning was in 1952, when Arthur Samuel wrote a computer program to play checkers. The program improved the more it played by studying which moves resulted in winning strategies.[6] Since then, machine learning has been used in a variety of different fields. Some examples include breast cancer detection, autonomous vehicles, and providing movie recommendations online, actions that were previously accomplished by human judgement. As machines begin to make more decisions for human individuals, it is important to ensure that the machines are safe and free of harmful biases. If machine learning technologies are trained from data that carry harmful biases, the algorithm itself will mirror those biases. For example, if the training data contains sexist associations of "woman" and "homemaker," the algorithms will consequently learn to be sexist. There are many inequalities in different societies due to factors such as wealth, race, and gender; these inequalities may be deepened thanks to biased algorithms.[7]

One example of how algorithms can be negatively trained is Microsoft's chatbot, Tay, named for the acronym "thinking about you." The conversational chatbot engaged with different tweets, and the more people tweeted at it, the smarter Tay became. However, some of the tweets Tay was using to learn how to communicate contained racist and sexist content. This tainted content made Tay learn to tweet racially charged tweets such as, "@NYCiticen07 I fucking hate feminists and they should all die and burn in hell," and "@brightonus33 Hitler was right I hate the jews."[8] Needless to say, Tay was terminated quickly after the initial implementation. While Tay's tweets were inconsequential to society, they are cause for concern. How can engineers account for skewed data in machine learning devices?

Tay is not the only example of algorithms mimicking the worst tendencies of people. In 2018, Amazon was forced to shut down an AI hiring tool that was used to scan resumes. The tool was trained based on the resumes of current Amazon employees, and learned to discriminate against women by penalizing resumes that included words like "Society of Women Engineers."[9] Professor Safia Noble illustrates another example of harmful machine learning in her book *Algorithms of Oppression*. There is evidence of Yelp hurting African American owned businesses by prioritizing the reviews of white middle-class users.[7] Furthermore, Dr. Aylin Calaskin has presented empirical evidence that machine learning picks up on human biases.[10] These examples demonstrate that not only is machine learning imperfect, it can also be harmful to marginalized groups like African Americans and women.

Average citizens are using artificial intelligence and machine learning to make decisions that were previously made solely by human judgement. The inequality of AI could become a human rights issue in the future if the engineering design process is not ethically managed. An example is Canada's backlogged immigration system. To mediate this issue, the Canadian government introduced algorithms in 2014 to automate the decision of whether someone should be allowed into the country. The algorithm flags markers such as whether a marriage is "genuine" or if someone is a "risk."[11] The use of this technology places marginalized communities at the mercy of machine learning, which could violate their human rights. Similarly, the state of California signed a law earlier this year eliminating cash bails in order to treat rich and poor individuals the same. Instead, algorithms will generate a score predicting the likelihood of re-arrest. Low-risk people will be released, and high-risk people will be placed in jail. The machine learning algorithms make decisions by comparing the age and previous charges to other similar criminals, but most of the factors that go into this algorithm are classified.[12] It is likely that groups that were harmed by the cash bail law, mostly people of color, could be directly harmed by the new law. Both the example of the Canadian immigration and the California law show that algorithms could either dramatically help or hurt marginalized groups, like refugees and individuals charged with crimes.

## Analysis

Who creates machine learning technologies? Apple was one of the first companies that released diversity reports of their employees. In 2014, "it was 70 percent male globally, and 80 percent male in technical roles. Two years later, in 2016, it was still 68 percent male globally, and 77 percent male in technical roles."[13] In the same year, 2016, Google released diversity reports indicating that "technical employees were 81 percent male."[13] At Airbnb, "10 percent of staff came from 'underrepresented groups' in 2016 (which means neither white nor Asian, the two groups that are well represented in tech companies) – but in technical roles, that number was only 5 percent."[13] While it has been a few years since these statistics were released, the trend remains the same: employees in leadership roles and technical positions are still predominantly white males. Does diversity account for a limitation in the engineering design process?

*How STS can help engineering design. Work in progress.*

## Conclusion

*Work in progress.*

## References

1 Plenke, Max. "We Figured Out Why Some Electronics Don't Work For Black People." (n.d.). Retrieved November 13, 2018, from https://mic.com/articles/124899/the-reason-this-racist-soap-dispenser-doesn-t-work-on-black-skin

2 Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science, 14(3)*, 399-441. Retrieved from http://www.jstor.org/stable/285355

3 Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences 4(2)*, 155-169. doi:10.1007/BF01405730

4 Zellner, M., & Campbell, S. (2015). Planning for deep-rooted problems: What can we learn from aligning complex systems and wicked problems?. *Planning Theory & Practice*, *16*, 457-478. doi:10.1080/14649357.2015.1084360.

5 Angra, S. & Ahuja, S. (2017). Machine learning and its applications: A review. *IEEE*. Retrieved from https://ieeexplore.ieee.org/document/8070809

6 Marr, B. (2016, Feb 19). *A short history of machine learning – every manager should read.* Retrieved from https://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning-every-manager-should-read/#6059b6d015e7

7 Noble, S. U. (2018). *Algorithms of Oppression*. New York, NY: New York University Press.

8 Vincent, J. (2016, March). *Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day*. Retrieved from https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

9 Smith, A. (2018, Oct). Amazon Shut Down Recruiting AI for Sexist Results. *PCMag.* Retrieved from https://www.pcmag.com/news/364315/amazon-shut-down-recruiting-ai-for-sexist-results

10 Calaskin, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183-186.

11 Molnar, P. & Gill, L. (2018). Bots at the gate: A human rights analysis of automated decision-making in Canada's immigration and refugee system. *The Citizen Lab*. Retrieved from https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf

12 Raphling, J. (2018, Sept 24). *California ended cash bail – but may have replaced it with something even worse*. Retrieved from https://www.thenation.com/article/california-ended-cash-bail-but-may-have-replaced-it-with-something-even-worse/

13 Wachter-Boettcher, Sara (2017). *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech*. New York, NY: W.W. Norton & Company.

## Bryn Seabrook

Bryn is an Assistant Professor in Science, Technology, and Society at the University of Virginia. Her research interests include bioethics, public participation in environmental policymaking, energy efficiency, climate change, negotiating the environmental-consumer nexus, analyzing American consumer culture, engineering education, engineering responsibility, and disability studies. Membership/leadership: American Society for Engineering Education, Society for Social Studies of Science (4S), STGlobal.