Investigation of Social Pressures on the Evolution of Machine Learning and Autonomous Robot Systems

STS Research Paper

Presented to the Faculty of the

School of Engineering and Applied Science

University of Virginia

By

Gabriel Argush

April 20th, 2020

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed: ____

Approved: ______________________________________ Date ________________________

Rider Foley, Department of Engineering and Society

General Research Problem

Autonomous robots and vehicles have been used in a wide range of applications for decades, both tangentially and separately from humans depending on the application. One relatively recent use case of these machines is in search and rescue missions, in which targets, which are usually victims of a disaster scenario, must be identified and evacuated to safety. To do this, autonomous vehicles must be capable of area mapping, target identification, and quick mobility. Because the "success [of an autonomous robot] is highly dependent on the time needed for first responder awareness that a human life is in imminent danger," it is imperative that the robot performs these actions consistently well (Lygouras, 2019, p. 1).

As our understanding of robotic autonomy grows every year, engineers develop more efficient ways to program and design machines for specific purposes. A popular trend nowadays is to implement machine learning algorithms to combat some of the complex problems faced in certain scenarios. However, as the complexity of robot systems and machine learning applications surges, so does the disconnect between developers of the technology and the general public. Many of those not involved in the development of autonomous robots see the technology as a sort of "black box" which is beyond their understanding. This, coupled with the perceived loosening of human control over robots over time, stirs up a general mistrust of machines to perform critical tasks. Current robot system designers and engineers must take into account the prevailing "measure of negative [public] attitude toward robots" to address the concerns of the general populace (Liang & Lee, 2017, p. 2). For the purposes of this prospectus, any reference of the general public refers to those who have been or are expected to be impacted by autonomous robots, but have very little to no contribution to the development of the robots. This includes previous victims and families of victims who have interacted with an autonomous robot, and those who may not have interacted with them but are concerned, at some level, with the application of robots as a means of saving people.

As we designed a 3D-printed autonomous robot to perform tasks such as indoor mapping, waypoint control, and target identification, my Capstone team at the University of Virginia aimed to acknowledge the strengths and challenges of the technology which affect our system. I intended to extend my team's work by researching the state of autonomous robot technology in society, in order to determine how such striking differences in perception of emerging technology came to exist. This thesis will combine the results of these two efforts to understand how faith in autonomous robots is shaped by differing views of their uses, effects, and risks.

Investigation of Social Pressures on the Evolution of Machine Learning and Autonomous Robot Systems

Engineers and developers have championed machine-learning algorithms for processing information in situations in which human brainpower would simply not be powerful enough. This is definitely the case in the realm of autonomous robotics, which are usually used out of a need to perform tasks which humans are unwilling or unable to do themselves. Failure during a search and rescue mission could result in injury or death, and would also damage the reputations of the engineers involved. The results of this study may help to better understand how people and the technology interact as a linked sociotechnical system, and why different groups, with inequitable levels of contribution to the system, react differently to how the technology is used.

Some sources have cited a profound fear in people of interacting with certain advanced robots, which may or may not have to do with the general population's degree of understanding how the technology actually works (Liang & Lee, 2017, p. 1). This fear, if it exists, would not be consistent across the relevant groups in the sociotechnical system, which may have to do with a disconnect in information between the public and higher technology. I hope to reach some conclusions concerning the non-engineers' motivations and priorities regarding autonomous robots, and would like to understand if it really is fear which causes the disconnect, or if there are other motives not often talked about in past research.

The initial step of my research is to identify what each stakeholder group considers to be the strengths, challenges, opportunities, and threats of robots and their effects, consolidating the totality of responses into a SCOT analysis. I expect to collect more data from those who would be considered to be in the non-engineer group, so my experiences with autonomous robotics will substantiate the other group's input.

A discussion on how to contextualize my research question from an ethical standpoint is the logical next step. Progress toward a sustainable sociotechnical system with robot technology becoming more advanced every year must consider the ethical implications of doing so. If robots are approaching a level of autonomy close to that in humans, and many humans feel uncomfortable with many aspects of the technology which may or may not be addressed in development, the disconnect between the stakeholder groups will only grow. I utilize a comparable example and its respective ethical framework to highlight some steps engineers have taken in other areas of applied technological design.

Case Context: Use Cases for Autonomous Robots in Society

My Capstone group's technical work served as relevant background experience for this paper, as it revolved around the development and testing of a 3D-printed autonomous robot whose core functions were to identify targets in an indoor area, reach target points and "return to base" when instructed to do so, as well as any other functionalities we planned to include over the course of the project. The MITRE Corporation sponsored our project and supplied us with the robot and the guidance on how to best implement the desired objectives. Many of these functions are critical in most applications of autonomous technology in society, whether it be for self-driving cars or for rescue and relief efforts. For the latter, of which I will concentrate most of my research effort, several conditions must be tested and met to minimize risk and avoid mistakes.

As such, a major focus of autonomous robot development today is to design a system which can overcome obstacles seen in general applications: physical and environmental hindrances, time constraints, and hardware recovery/ repair problems. For example, we programmed the robot to perform functions which are likely to be applicable in scenarios in which humans are not able to operate on their own, due to constraints such as time or small spaces. A robot system like ours would also have the benefit of being easily constructed and repaired in the field, without the need for stocking a large inventory of parts. In addition, by giving operational personnel access to the design, the systems can be modified in the field to meet unforeseen mission requirements, essentially enlisting the operators as "hackers" to improve the robot's design.

As one can imagine, there are countless possible sources of failure in each design choice, functionality, and method of execution when working with intricate, multicomponent machines. My team's work was no exception, and has such discovered several inconsistencies between what we planned for the robot to do, and what we observed in action. For example, when approaching a hard surface, we believed the robot had been programmed to stop a set distance away from the surface, and not continue forward even when instructed to do so. The robot did this, but we also observed that the robot would not allow for backward movement either- it had to turn left or right at first to escape, which is not optimal. On the hardware side, the current iteration of the robot uses a light-based sensor and a robust algorithm for positioning and mapping, but we often encountered inconsistencies in the data procured. It was our responsibility to iteratively improve the system based on defined requirements and goals, to ensure these discrepancies did not occur in the future. There is no doubt that correcting malfunctions and inaccuracies are at the forefront of many engineering teams' lists of priorities, as it was for ours.

My team's deliverable was a robot with enhanced sensing and communication designs, which better performs under the determined criteria for success. Our team aimed to test the current system's capabilities, in order to assess the performance of the current and future prototypes as they developed. In order to do this, we needed to explore relevant open-source frameworks to improve the system's ability to create maps of unknown spaces and provide robust control with low latency. It was found to be important to allow an operator to drive the robot using a first-person view setup to be able to tag objects of interest from their perspective. All in all, the team made it a goal to iteratively modify the robot and system to better meet concrete objectives for the project, as well as collect quantitative and visual data on robot performance as necessary.

Research Question and Methods

In this paper, I proceed to discover each stakeholder group's perspective regarding the use of autonomous robot technology in society. I choose to focus on two major groups for my analysis: the engineers programming autonomous robots to do the "saving," and the victims who are being saved in some relief scenario. My primary analysis is on autonomous robots used in rescue and relief efforts, in which targets- usually victims of a disaster scenario- must be identified and evacuated to safety.

Research Question

This thesis answers the question, how have conflicting societal perspectives on robot technology have contributed to the differences among these stakeholders regarding faith in autonomous robots and interests in their uses?

Methods and Data Sources

My principal data collection tool for research was that of a survey. The survey asked a multitude of questions regarding the respondent's initial thoughts about machine learning and robot technology, comfortability in using human-controlled vs. fully autonomous technology in critical scenarios, and biggest fears regarding the employment of nonhuman decision-makers. I did not strictly confine the sampling distribution in terms of demographics or understanding of the technology, as I wanted to collect perspectives and comments without bias of age, gender, interest in STEM, etc. Overall, I collected 74 responses, most of which have answered every question asked in the survey. Analysis of responses were done using Minitab to make use of robust statistical testing and modeling techniques which are applicable to this study. Nonparametric models were used when distributions could not reasonably be assumed to be Normal, and models used a significance level of 0.05 unless otherwise stated. A more detailed description of the questions asked in the survey can be found in the Appendix.

The majority of respondents to the survey are from 18 to 28 years old. Of the 69 who indicated age, 56 belong to this group (81%). 28 respondents are of undergraduate university age (18-21, 41%). Gender data ended up being a bit skewed, with 70% of respondents reporting to be female, and 29% responding male. The ethnicity breakdown of the received 73 responses is as follows: 51% Caucasian, 15% Asian, 15% White not Caucasian, 7% Black/African, 6% Hispanic/Latinx, and 6% wished not to respond. For highest level of education received or in process of receiving, 55% reported a Bachelor's degree, 22% a Master's degree, 12% a high school diploma or equivalent, 7% an Associate's degree, and 4% below high school degree. Interest in STEM subjects was asked to be rated on a scale from 1 to 4, 4 being of highest interest, and had the following breakdown: 15% responded with 1, 36% with 2, 29% with 3, and 19% with 4.

Results

I divide the analysis into three sections for analysis: general outlook, a use case scenario, and comments on fears and challenges. The findings in each section provide evidence of priorities and preferences across several demographic variables which differ from those that engineers usually consider most important, such as "path planning, resource allocation, and motion planning" (Wurman et al., 2008, p. 14). Overall, it appears the disconnect between the two stakeholder groups exists due to different levels of comfort with robots, which is based on the degree of human control over the them. This is something engineers must address with suitable ethical standards if the technology is to be used in society.

General Exposure and Outlook

Most respondents have heard of autonomous robots being used in certain scenarios, but have not personally seen them. Of all 74 responses, 92% are familiar with the technology in self-driving cars, and 40.5% know of both grocery stores and delivery services employing the autonomous vehicle technology, but only 27% have heard of the robots being used in rescue/relief missions, the focus of my research. Surprisingly, of 63 recorded responses, only 12 (19%) have personally experienced an autonomous vehicle in some use capacity.

Respondents more or less have similar outlooks on the effects artificial intelligence and autonomous robots will have in the future, even though they are generally used in different applications. As similarly found in Liang & Lee's analysis of societal technological fear, the results here "support the contention that people do respond to autonomous robots and artificial intelligences in an empirically indistinguishable way"; on a scale from 1 to 5, 5 being most positive, 44% and 45% reported the mode of 4, and the distributions of responses do not have statistically different means (p=0.760) (2017, p. 379). Responses of outlook were fitted to a generalized linear model across all demographic variables. Major factors which influenced outlook positively are: Asian ethnicity (p=0.018) and Master's degree (p=0.05); likewise, factors which lowered outlook responses are: Female gender (p=0.011), and arguably belonging to the highest age group (p=0.13). Having no higher education had a statistically significant negative contribution, but has been excluded due to very low group size. Interest in STEM did not have a substantial affect, nor did other age groups, ethnicities, or education levels.

Use Case Scenario

The next section of the survey included a short story about a team of engineers who wish to test their new prototype robot's ability to "detect human survivors and navigate a space in a destroyed building" scenario. Respondents were asked to rate their comfortability (1-5, 5 highest) with the test if the robot were to be used in human-controlled and fully-autonomous modes, as well as if the tests were to take place in the years 2020 and 2040.

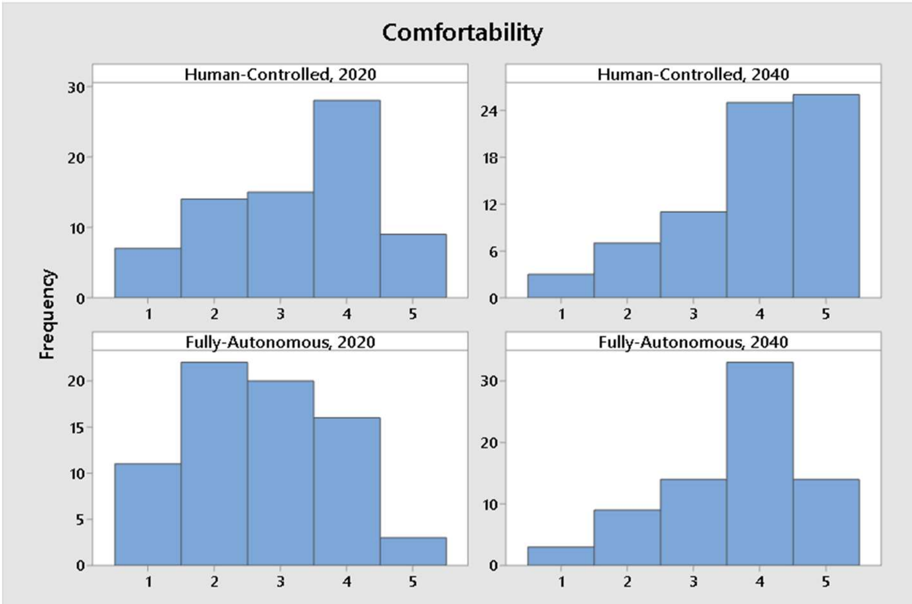

Figure 1: Histograms of comfortability with certain robots; human-controlled vs. fully-autonomous, 2020 vs. 2040 (Argush, 2020)

Respondents appear to be somewhat comfortable with a human-controlled rover in 2020, but are much more inclined to trust one in 2040, as 51 of 72 (71%) responded as Comfortable or Very Comfortable for a test in the future. Responses were overly negative for a fully-autonomous test in 2020, but the distribution of responses for an autonomous test in 2040 shows more trust in the technology. Few factors were statistically significant in models for each of the four distributions, and so will not be discussed.

The next section asked to rate the following five common objectives in autonomous robot systems by importance: "High target identification accuracy"; "Move and act quickly"; "Try to save as many lives as possible"; "Be able to work in a new environment"; "Intuitive operation and maintenance". "Try to save as many lives as possible" was overwhelmingly the most popular with 44 of 74 (59%) votes, and "High target identification" was second highest with 20 of 74 (27%), to constitute 86% of total responses for the most important objective. "Intuitive operation and maintenance" was rated the least important choice for 42 of 74 (57%), and "Be able to work in a new environment" was rated the most for 4th-most important with 36 of 74 votes (49%). It is noteworthy that the objectives which would most concern an engineer in charge of an autonomous robot's operation, such as the ability to work in a new environment or to allow for ease of maintenance, were the predominantly the least important for those taking the survey.

	1st Choice	2nd Choice	3rd Choice	4th Choice	5th Choice
High target identification accuracy	44	19	18	10	7
Move and act quickly	7	26	22	12	7
Try to save as many lives as possible	20	15	11	3	1
Be able to work in a new environment	2	9	10	36	17
Intuitive operation and maintenance	1	5	13	13	42

Figure 2: Ranking of robot objectives by tally (Argush, 2020)

Comments on Fears and Challenges

Respondents were asked to comment on which issues struck them the most with regard to both the short story example and in general. These comments, 58 in total, were then condensed into nine major categories and tallied, as shown in the below table. Some comments also indicated some benefits of the technology from the responder's viewpoint, which will be discussed below. I explicitly structured these questions to assess which factors are perceived to be the most threatening or challenging from the perspective of someone who might actually interact with an autonomous robot, as well as reveal any prevailing positive sentiments within the group. The answers were then inputted into a SCOT analysis matrix for analysis.

Strengths	**Challenges**
Helpful for modern life	Malfunctions (19)
Useful for those in need of extra assistance	Fails at something human would succeed in (8)
Useful for dangerous, life-threatening jobs	Makes things worse (7)
	Technical limitations (7)
	Safety (3)
	Environmental obstruction (2)
Opportunities	**Threats**
Future developments in the technology	Human displacement (9)
Widespread publicity of success	Corporatization (2)
	Hacking (1)

Many of the major challenges reported dealt with a robot's failure to do its job correctly. More broadly, the biggest pain points in respondents' minds were potential malfunctions/ glitches/ errors which could have been avoided if a human did the work or with more accurate technology. Some went further and specified large-scale problems such as the displacement of human work and use of robots for political or nefarious gain. My hypothesis for this section was that those more interested in STEM were likely to make comments regarding macroethical and technological threats more so than malfunctions and disaster-oriented mission results (Jorge et al., 2019, p. 30). However, interest in STEM had very little to do with the type of comment made, as I had originally anticipated. Likewise, several respondents noted strengths and opportunities for advancement using robot technology with which engineers would similarly respond, such as improving quality of life and performing dangerous tasks.

Discussion

Ethics of care

The logical first step in addressing the growing ethical concerns of the technology is to use an applicable ethical framework. It is clear from the survey data that non-engineers are guided by principles of care which engineers tend not to put at the forefront when designing a robot. Those who will be the ones to interact with an autonomous robot in the field are likely to consider the importance of technical capabilities, but are much more concerned with their own safety, as well as the chance the robot makes a mistake and makes the situation worse.

Respondents reported that they would be much more comfortable with a semi-autonomous robot than a fully-autonomous one, and expect the latter to perform tasks more accurately and safely due to advancements in technology down the road. A plausible method of envisioning how to apply an ethical baseline to a technology as it develops over time could be to attribute a human code of ethics to the robot's operation, effectively "[assigning] responsibility to the robot itself" as humans would to other humans (Lin et al., 2011). Doing so would affix a measure of accountability which increases as robots become more autonomously independent over time. This would also be consistent with the "social blame hierarchy" hypothesis, in that "people assign blame on the basis of a hierarchy that is based on perceived autonomy"; in essence, if robots become more competent autonomously as they develop, a higher level of scrutiny should be applied to ensure the robot meets the criteria of the ethical code of care (Furlough et al., 2019). Such a hypothesis is further supported by my survey data if a connection between attribution of blame and trust in technology can be established. As evidenced, fully-autonomous robots are trusted far less than ones operated by humans are for an identical task, and humans alone would be trusted more than both, thus forming a hierarchy based on human control.

In short, as robots become more capable as moral agents and humans "expect [them] to bear the responsibility for their actions and make ethical decisions, we may need to be prepared to cede more control to them," and, consequentially, more blame as well (Gerdes & Thornton, 2015). Although engineers have much work to do before reaching a stage in which humans perceive themselves and robots at the same hierarchical level of blame, this line of thinking could more closely align robot operations with acceptable ethics of care for victims.

Autonomous robots in other fields

Some of the aspects regarding the ethics of care can be seen in autonomous robot systems designed to assist the elderly. One such example can be seen in the emotional support robots which take the form of dogs, seals, and koalas. These autonomous robots are designed to respond to sensory information as a pet would: they can change responses based on how they are touched, maintain eye contact toward sources of sound, and repeating previous actions which garnered praise from the owner. Paro, a robotic seal created in Japan, can even go so far as to analyze symptoms of certain disorders and change behavior to reduce stress for patients, allowing for "greater independence for those with dementia or other aging brain systems" (Lin et al., 2011, p. 268). Although the objectives of emotional support robots and autonomous robots used in the field differ, and the former are "clearly not themselves moral agents" as a robot identifying victims of a disaster might be, the reasons engineers designed the former in the way they did could be illustrative as an example for others (Calo et al., 2011, p. 21). A goal-oriented and consumer-centric approach, involving a deeper understanding of all stakeholder needs, would be more aligned with appropriate care ethics.

Limitations & caveats

In consideration of the limitations of my research, it would be rash to ignore the downsides of administering a relatively unstructured survey. In an ideal setting, the survey would be sent out or physically given to a pool of respondents in a more controlled way. This would minimize potential survey biases which could otherwise be out of my control, such as neutral responding and acquiescence biases, which are typically represented when the person is unsure or not fully informed about the subject matter. If possible, a larger group of respondents with more representative characteristics of the general population would provide a much better picture of the intended survey landscape.

Future considerations

If I were to perform similar research in the future, I would alter some aspects of my process. First, I would revise my survey to be a bit shorter, and utilize fewer open-ended questions for the respondent to answer. I found that several people filled in the multiple choice and slider-style questions, but didn't respond to the qualitative questions aimed to gauge their fears and comments. This is likely due to the desire to finish the survey as quickly as possible, which is understandable. I think forcing the person to answer all questions might be too harsh, given they are tasked to read a short story in supplement to answering the questions. Trimming the fat of the survey may likely yield higher overall response rates for every question. I would also broaden my survey population by reaching out to even more groups of people who might be willing to assist my research. My sample size ended up smaller than I had intended, which only pushed my margin of error up; aiming for 150 to 200 responses would have removed substantial error in my analysis (Bartlett et al., 2001, p. 47-48).

Engineering impact

The implications of my research reinforce some of the engineering principles I've learned in a more practical setting. The crux of systems engineering is being able to make data-driven decisions for others based on their requirements and desired objectives. One of the most effective ways to best achieve someone's goals is through "value-sensitive design," to gain a deeper understanding of their intentions and which factors and values cause them the most worry (van Wynsberghe, 2013, p.410). By being able to visualize and analyze the responses by people from various walks of life, I could better understand how to view my year's work on robots as a "sociotechnical engineer" of sorts- certainly better than I had as just a technical engineer- and not solely be motivated by optimization functions and applying technical know-how.

Conclusion and Next Steps

It is clear that potential stakeholders who are the ones to interact with autonomous vehicles have very different priorities for the robots than what engineers initially have in mind. Under the current paradigm in which robots get developed, engineers do not consider the input from those receiving care as much as they should; otherwise, there wouldn't exist such a disconnect in the two groups' priorities. Perhaps the respondents to my survey believe that in addition to technical advancements in the relevant fields of autonomous robot technology, they also hope that engineers in the future will better address the qualms people have with allowing nonhuman entities to make critical decisions for humans.

It would be valuable to take note of how autonomous robot technology is used to great effect in the healthcare field. Not all nonhuman technology has to appear cold and heartless; a greater focus on meeting the requirements of those who interact with robots in the field would make strides toward finding a common ground for future work. Development in machine learning and autonomous technology is unavoidable in the coming years. It only seems appropriate that this growth should be deliberate in its goal to impact all stakeholders equitably.

References

Bartlett, J. E., Kotrlik, J. W., & Higgins, C. C. (2001). *Organizational Research: Determining Appropriate Sample Size in Survey Research*. 8.

Calo, C. J., Hunt-Bull, N., Lewis, L., & Metzler, T. (2011, August 24). Ethical Implications of Using the Paro Robot, with a Focus on Dementia Patient Care. *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*. Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence. https://www.aaai.org/ocs/index.php/WS/AAAIW11/paper/view/3808

Furlough, C., Stokes, T., & Gillan, D. J. (2019). Attributing Blame to Robots: I. The Influence of Robot Autonomy. *Human Factors*, 18720819880641. https://doi.org/10.1177/0018720819880641

Gerdes, J. C., & Thornton, S. M. (2015). Implementable Ethics for Autonomous Vehicles. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte* (pp. 87–102). https://doi.org/10.1007/978-3-662-45854-9_5

Jorge, V., Granada, R., Maidana, R., Jurak, D., Heck, G., Negreiros, A., … Amory, A. (2019). A Survey on Unmanned Surface Vehicles for Disaster Robotics: Main Challenges and Directions. *Sensors, 19*(3), 702. https://doi.org/10.3390/s19030702

Liang, Y., & Lee, S. A. (2017). Fear of Autonomous Robots and Artificial Intelligence: Evidence from National Representative Data with Probability Sampling. *International Journal of Social Robotics*, *9*(3), 379–384. https://doi.org/10.1007/s12369-017-0401-3

Lin, P., Abney, K., & Bekey, G. A. (2011). *Robot Ethics: The Ethical and Social Implications of Robotics*. MIT Press.

Lygouras, E., Santavas, N., Taitzoglou, A., Tarchanidis, K., Mitropoulos, A., & Gasteratos, A. (2019). Unsupervised Human Detection with an Embedded Vision System on a Fully Autonomous UAV for Search and Rescue Operations. *Sensors*, *19*(16), 3542. https://doi.org/10.3390/s19163542

van Wynsberghe, A. (2013). Designing Robots for Care: Care Centered Value-Sensitive Design. *Science and Engineering Ethics*, *19*(2), 407–433. https://doi.org/10.1007/s11948-011-9343-6

Wurman, P. R., D'Andrea, R., & Mountz, M. (2008). Coordinating Hundreds of Cooperative, Autonomous Vehicles in Warehouses. *AI Magazine, 29*(1), 9–9. https://doi.org/10.1609/aimag.v29i1.2082

Appendix

Section	Question	Answer Choices
Background Questions	In which cases have you heard of autonomous vehicles/robots being used?	Multiple Select • Self-driving cars • Rescue of relief missions • Grocery stores • Delivery services
	Have you personally encountered an autonomous vehicle in use? If so, where?	Yes or No
	How positive is your outlook on the effects of autonomous robots in the future, in cases involving human lives?	Scale from 1: Very Pessimistic to 5: Very Optimistic
	How positive is your outlook on the effects of machine learning/ artificial intelligence in the future, dealing with humans and everyday life?	Scale from 1: Very Pessimistic to 5: Very Optimistic
Short Story	How comfortable would you be if a human-controlled robot were tasked with navigating the space and identifying victims?	Scale from Very Uncomfortable to Very Comfortable, for both "In 2020" and "In the future (say, 2040)"
	How comfortable would you be if a fully-autonomous robot were tasked with navigating the space and identifying victims?	Scale from Very Uncomfortable to Very Comfortable, for both "In 2020" and "In the future (say, 2040)"
	What would you say is your biggest fear regarding the use of a fully-autonomous robot (something about the scenario, technical capabilities of the robot, etc.)?	Open-ended short answer
	How important would you rank each of the following objectives regarding the robot?	Choosing 1st through 5th choices of the following: High target identification accuracy; Move and act quickly; Try and save as many lives as possible; Be able to work in a new environment; Intuitive operation and maintenance
Demographic Questions	Age	Short response
	Gender	Male; Female; Choose not to identify; Other
	Ethnicity	Asian; Black/African; Caucasian; Hispanic/Latinx; Native American; Pacific Islander; White not Caucasian; Prefer not to answer
	Highest level of education	No schooling completed; High school degree or equivalent; Associate's degree; Bachelor's degree; Master's degree; Doctoral degree
	Interest in STEM	Scale from 1: No interest to 4: High interest

Appendix A: Survey questions and answer choices, condensed