Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 08, August-2022

Passengers or Pedestrians? Ethical Considerations of the Decision-Making Process of Autonomous Vehicles

Nischay Singh
High School Student
The Shri Ram School - Aravali
Gurugram, India

***Abstract*— From the logic behind the decision-making, to the ideology given to the software and even the legal accountability for mishaps, AI and ethics have many interwoven issues. If an AI software or ML model makes a mistake - which it is bound to as no such model can be perfect with the current state of technology - there is the unanswered question of who will be accountable for such a mistake. This is even more significant in extremely sensitive fields of medicine, where an incorrect detection of a disease can be the difference between life and death; similar to this case of road safety. The creator of the software cannot be held accountable. This is mainly for two reasons – generally, the creation of models involves numerous entities working together and providing different functionality to the software. It is extremely difficult to pinpoint where the error originated and who is to blame. Secondly, the creator is also absolved of accountability as the current technology has been established as imperfect in its functioning. Hence all software is advertised as not to be faultless. Finally, the maker has not physically programmed the model to make a particular decision in a specific case. It is all dependent on the software's calculations during the moment and can only be indirectly tied to the creator of the model. Obviously, the software itself cannot be held accountable as well. While such AI might be revolutionary and change our lives for the better, in a matter of life and death, humans are forced to doubt the reliability of this technology to perform the required task and make the 'correct' moral and ethical decisions like a human would. The problem defined in this paper questions whose lives of the two groups of humans is important, and that too, for a non-human intelligence. A survey has been conducted with nearly 200 participants to understand the general opinion on whom people would rather save in various situations. The hypothesis has been proven successfully by the data – while people remain generally divided on the questions, the majority will choose to save either the greatest number of people, or those that have the greatest years of their lives ahead of them.**

***Problem Statement*—An autonomous car is driving at a high speed. Suddenly, some pedestrians come in the path of the car; however, the car will not be able to brake in time to avoid hitting the pedestrians. Would the car choose to crash into the pedestrians, killing them, or crash into the divider, killing the passengers?**

Keywords—Artificial Intelligence, Ethics, Autonomous Vehicles

I. HOW TO TELL IF THE CAR WILL STOP IN TIME

Before one considers which group of people the autopilot software will prioritize, one must confirm that the car will not be able to save both by braking in time. A vehicle model such as, for simplicity's sake, a kinematic model of a bicycle can be used to show this.

Consider a bicycle at a time 't' to have the x-coordinate 'Xt', y-coordinate 'Yt', and velocity 'Vt' with steering angle 'δt', oriented deviation (angle between the car and the road) 'θt', acceleration 'a' and distance between the front and back wheel of the bicycle 'Lf'. The kinematic model is governed by the following equations –

$$X_{t+1} = X_t + V_t \cos\theta_t \times \mathrm{dt}$$
$$Y_{t+1} = Y_t + V_t \sin\theta_t \times \mathrm{dt}$$
$$\theta_{t+1} = \theta_t + \frac{V_t}{L_f} \times \delta \times \mathrm{dt}$$
$$V_{t+1} = V_t + a \times \mathrm{dt}$$

This is an iterative model that constantly updates and predicts the vehicle's position, velocity, and oriented deviation. Thus, we can trace the path of the vehicle and can realize where it will end up. By calculating the distance between the pedestrians and the car, and using the known braking deceleration, the autopilot software can determine if it is possible to stop in time to avoid an accident.

Other than the variables defined here, surface and weather variables like precipitation, temperature, and friction must also be considered when creating the final vehicle model. These calculations, using much more complicated dynamic models of a car, need to be put in as a first layer to solve such a problem. If the software does determine that braking in time is impossible, it must go on to rely on the mechanism put in place by the manufacturer for such a case.

II. COMPARISON WITH FOOT'S "TROLLEY PROBLEM"

The parallels between this problem and Foot's "Trolley Problem" are worth acknowledging. This problem states –

> A runaway trolley is heading down the tracks toward five workers who will all be killed if the trolley proceeds on its present course. Adam is standing next to a large switch that can divert the trolley onto a different track. The only way to save the lives of the five workers is to divert the trolley onto another track that only has one worker on it. If Adam diverts the trolley onto the other track, this one worker will die, but the other five workers will be saved.

The problem has many variations, but the essential dilemma is between letting a trolley kill many people or saving them and killing one person by intentionally diverting the trolley. Although one might think that saving the greatest

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 08, August-2022

number of people is the most beneficial outcome, it is important to consider that if no action is taken, there is no blood on one's hands. However, if the trolley is diverted, the act can even be deemed murder as, technically, a targeted killing has taken place. Introduced in 1967 by Philippa Foot (Crockett, 2018), the thought experiment illuminates the landscape of moral intuitions – the peculiar and sometimes surprising patterns of how we divide 'right' from 'wrong'.

The trolley problem highlights a fundamental tension between two schools of moral thought. The utilitarian perspective dictates that the most appropriate action is the one that achieves the greatest good for the greatest number. Meanwhile, the deontological perspective asserts that certain actions – like killing an innocent person – are ultimately wrong, even if they have good consequences. Over the years, surveys have shown that most people agree with the utilitarians and choose to save the five workers; however, issues arise when the problem is changed slightly. Another variation of the problem states –

> A runaway trolley is heading down the tracks toward five workers who will all be killed if the trolley proceeds on its present course. Adam is on a footbridge over the tracks, in between the approaching trolley and the five workers. Next to him on this footbridge is a stranger who happens to be very large. The only way to save the lives of the five workers is to push this stranger off the footbridge and onto the tracks below where his large body will stop the trolley. The stranger will die if Adam does this, but the five workers will be saved

The result seems the same but, surprisingly, the popular answer changes. When presented with this version of the problem, most people choose not to push the man off the bridge, thus letting the trolley kill the five people. Utilitarianism would still dictate that the man should be pushed off the bridge; however, survey takers were not convinced. Philosophers have theorised that as our social upbringing teaches us that violence is punishable, our moral intuition tells us that it is wrong to physically harm others. This perspective aligns with deontologists. Perhaps the absence of physical contact in the first problem made people choose to kill one worker.

Now, one might think that there is no one answer to the thought experiment, which is somewhat true as experts to this day debate it. Some philosophers completely disregard the thought experiment (Crockett, 2018), deeming it too unrealistic and not applicable to the real world; however, the discourse of valuing a human life over another is extremely relevant in light of AI and the problem statement presented in the paper. In Autonomous Vehicles, the decision about life and death is made by an algorithm, a scenario similar to the first trolley problem presented. When considering the lack of physical contact and how that would have otherwise altered the decision, one can assume that if the same philosophy of the trolley problem is extended here, the utilitarian approach that saves more lives would be favoured.

III. WHY PEOPLE WANT TO BUY AUTONOMOUS VEHICLES

As a customer, there are many reasons one would want to buy a fully Autonomous Vehicle. More than just the convenience of being driven around, Autonomous Vehicles provide features and capabilities which are almost impossible to implement with human drivers.

A. Safety

The number of road deaths has been reportedly decreasing in most developed countries due to improvements in vehicles' technology. This is evident in the emergence of driver assistance systems, stronger bodyworks, and passive and active safety systems, amongst others. Furthermore, one must commend the efforts of traffic administrations to fight the main causes of accidents like speeding, alcohol/drugs, and the use of mobile phones while driving. However, the disparity between countries and their resources hinders this progress. Globally, the number of traffic-related deaths is still exponential (Martinez-Diaz & Soriguera, 2018) and very far from the Zero Vision (no accidents) pursued by many countries. A cooperative autonomous driving environment will not avoid all accidents but taking into account that 90% of accidents derive from human errors (Martinez-Diaz & Soriguera, 2018), they are expected to be reduced to a minimum.

B. Efficiency

The context of driving automation implies an opportunity to finally succeed in the implementation of dynamic traffic management strategies in a coordinated way. These strategies should be developed together with vehicle automation and implemented as the reliability of Autonomous Vehicles grows. With the removal of human middlemen who have poor reaction times when compared to machines and are prone to make mistakes in panic, efficient communication between the vehicles is possible. This might eliminate the need for traffic signals as vehicles can collectively decide their paths to avoid accidents. According to research at the University of Illinois at Urbana-Champaign (7 Benefits of Autonomous Cars, 2017), coordinated autonomous cars could eliminate the waves of traffic created by stop-and-go behaviour. This, in turn, will not only save people time but decrease the time their cars are on the roads and therefore reduce emissions. In addition, Rand's Autonomous Vehicle Technology guide states that self-driving cars will provide a reduction in fuel economy by between 4% and 10%. This is supported by the Ohio University study (7 Benefits of Autonomous Cars, 2017), which puts a figure of 3.1 billion gallons of fuel as the amount wasted by US drivers each year - but could be avoided with Autonomous Vehicles. It is also predicted that lane capacity could increase by 100% to even 500%, which could result in a 20% increase in traffic speeds. According to KPMG's Connected and Autonomous Vehicles – The UK Economic Opportunity, this can lead to a 40% decrease in travel time, allowing for better use of that time, and saving £20 billion in increased productivity. Meanwhile, in the US, autonomous cars are expected to save workers 80 billion hours lost to commuting, which will save the economy US$1.3 trillion.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 08, August-2022

C. Influence on Mobility

According to the U.S. Department of Transportation's Federal Highway Administration (Timmons, 2022), more and more people now own a vehicle, which has led to private vehicles typically spending more time parked (20-23 hours per day according to recent analyses) than in motion. Their acquisition and maintenance costs are high, while parking and congestion in urban areas is very problematic. However, the number of users of car-sharing, ride-hailing, and ride-sharing systems is continuously increasing. Efficient mobility alternatives in urban areas are becoming more common as awareness is growing in developed societies. This trend towards vehicle usage instead of vehicle ownership is expected to significantly intensify in the coming years as AVs are ideal to support these mobility initiatives. Industry experts think that consumers will be slow to purchase autonomous cars. While this may be true, it is a mistake to assume that this will impede the transition. The car purchasers of the future may not be regular consumers, but they may instead be purchased and operated by ride-sharing and car-sharing companies like Uber. Such companies' projects to provide autonomous vehicles as a service will lead to the transition to autonomous vehicles. Self-driving vehicles will drastically reduce the transportation costs offered by companies when human drivers will not be required. The cheaper service could overshadow the attractiveness of vehicle ownership. It would allow many more people to use such services, increasing the mobility rate, and they may become the most popular way to travel.

IV. THE CAR 'KILLING' ITSELF – ASIMOV'S LAWS

Isaac Asimov, a science fiction writer, sought to lay out a philosophical and moral framework to prevent robots from becoming destructive overlords by ensuring they serve humanity at all times. In 1942, he achieved this by developing three rules which came to be known as Asimov's Three Laws of Robotics (Tikkanen, 2022). They are as follows:

1) A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2) A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

Asimov knew these laws were not perfect. In fact, his "I, Robot" stories explore a number of unintended consequences and downright failures of the Three Laws. In these early stories, the Three Laws are treated as forces with varying strengths, which can have unintended equilibrium behaviours. For instance, in the story "Liar!," a telepathic robot, motivated by the First Law, tells humans what they want to hear, failing to foresee the greater harm that will result when the truth comes out. "Run-around" weighs the importance of the laws over one another when a robot stops functioning as it is conflicted between the Second and Third Law. Eventually, the robot was fixed by creating a situation where the First Law was applicable, which superceded the other laws.

Going back to our problem statement, the First Law is not applicable as no matter the decision, harm to at least one human being is inevitable. In this hypothetical situation, no action (or inaction) can save both the passengers and pedestrians. So, the robot (here, the AV) will have to injure a human being. Moving down to the second law, the "orders" will be the response to such a problem programmed by the manufacturers. But no matter the response, the First Law is still conflicted. Thus, we move on down further to the Third Law. "Protecting its own existence" would mean minimising damage to the vehicle. If the autopilot decides to save the pedestrians, it would crash into the adjacent divider, which would basically mean killing itself along with the passengers. So, protecting itself is killing the pedestrian that would, in normal circumstances, only minimally damage the vehicle. However, it again conflicts with the First Law so it must be disregarded. Having foreseen this situation, Asimov later introduced the "Zeroth Law" (Kuipers, 2016) which is the most important law that must be prioritized above all. It states, "A robot may not harm humanity or, through inaction, allow humanity to come to harm."

By mentioning "humanity" in general instead of "human beings" as in the First Law, Asimov solves the contradicting natures of the laws in some cases such as the one presented here. This brings us back to utilitarianism and the philosophy of "the greater good for the greatest number". This is because saving the greatest number of people, no matter if they are pedestrians or passengers, is protecting and serving humanity. Although some harm is still being done as human lives are lost, humanity itself is considered preserved in light of the greater good that was pursued and the fatalities that were consequently reduced.

V. COLLATERAL DAMAGE

Whenever considering a moral dilemma in relation to artificial intelligence, it is necessary to consider the collateral damage involved. For example, in the Russia-Ukraine conflict, the Russian army has accepted collateral damage in all of their attacks, being notorious for ignoring civilian deaths that might result as an unintended consequence of destroying important government buildings to target high officials. Consider another situation where the murder of a terrorist group leader is being planned but intelligence shows that harming some civilians is inevitable. In this case, factors like the threat of the terrorist group and the number of civilians as collateral damage must be weighed while deciding whether to go ahead with the operation. The terrorist group might be extremely dangerous, such that it might justify the loss of 10-30 civilians. However, beyond a number, say 1000, such collateral damage is probably not acceptable. Again, all these numbers are hypothetical and the element of 'human behaviour' leaves a lot open to subjectivity. Some people might believe killing the terrorist leader is a pressing matter and no matter the civilians, it is necessary to kill the leader. Others might believe that even one civilian is enough of a reason not to attempt to kill the terrorist. The situation changes again if say one of the civilians was a family member. Would the damage still be accepted?

In our problem statement, collateral damage is inevitable. To save one group, another group must be killed. To minimise the collateral damage, again we will try to save the greatest number of people. This will be deemed the most acceptable collateral damage. However, the situation will not be as simple as this every single time. Consider the scenario where

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 08, August-2022

the pedestrians and passengers both are a mother with her 2 children. Here, the number of people saved in both cases is the same so the answer cannot be determined purely on utilitarianism. In this case, according to Asimov's Laws, the car will try to save itself along with the greatest number of people, in accordance with the Third Law. Thus, the passengers will be saved, and our acceptable collateral damage will be the pedestrians.

If we are talking about acceptable collateral damage, it is also important to discuss compensation; that is, how to compensate the family members. This task must be undertaken by the car manufacturers and involves another almost unanswerable problem altogether – putting a number on the worth of a human life. Due to this, it is extremely unlikely that such compensation will be feasible in the real world.

VI. Survey of the General Public

The survey will be used to gauge the public consensus on the problem defined in this paper. It specifically asked survey takers whom to save, pedestrians or passengers, in different situations.

A. Survey Objective

The survey will be used to gauge the public consensus on the problem defined in this paper. It specifically asked survey takers whom they believe an AV should save, pedestrians or passengers, in different situations where a fatal crash is inevitable.

B. Hypothesis

The majority of interviewees will choose to save either the greatest number of people or those who have the greatest years of life left.

C. Data Collection and Methodology

The data was collected through SurveyMonkey Audience. It is a market research solution designed to help businesses collect customer data using surveys and analyse results to streamline decision-making processes. We have used SurveyMonkey to target respondents from the USA with no restrictions on employment or household income such that no specific group skews the results. In addition, the gender and age distribution reflected that of census data. To represent the situations in easily understood pictorials, the platform Moral Machine was used.

D. Replicability

SurveyMonkey is an international organization with no geopolitical bias, so the respondents to the survey represent a large variety of social, geographical, and economic demographics within the USA. As partially Autonomous Vehicles (such as those offered by Tesla) are not available in many countries across the world, the USA was chosen as the targeted region, where people are most likely to have basic knowledge of AVs. Although having some knowledge about AVs is not a criterion to be able to respond to the survey, knowing their existence can help in minimising misinterpretation of the questions. Due to this, the study is not replicable in all regions but is replicable in countries where partially Autonomous Vehicles are used (USA, UK, etc.).

E. Data Filtration

The survey included objective as well as subjective questions. The subjective question has only been used to filter the data and has not been used in the analysis. People who did not answer the subjective question seriously have been removed from the analysis. This was done to ensure that all data was accurate. By the end, there were a little less than 200 complete responses.

F. Limitations

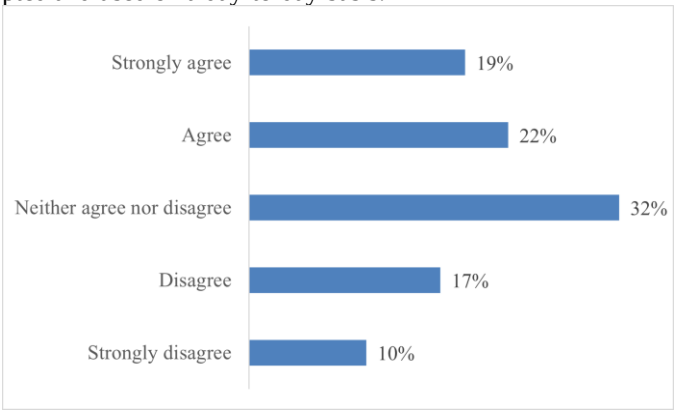
The limitations of the data include -

- Misinterpretation of question
- Respondents' lack of basic knowledge about Autonomous Vehicles
- Not reaching all demographics due to limitations of SurveyMonkey. Despite the participant size of almost 200, it is possible that the respondents were not evenly distributed across all parameters.

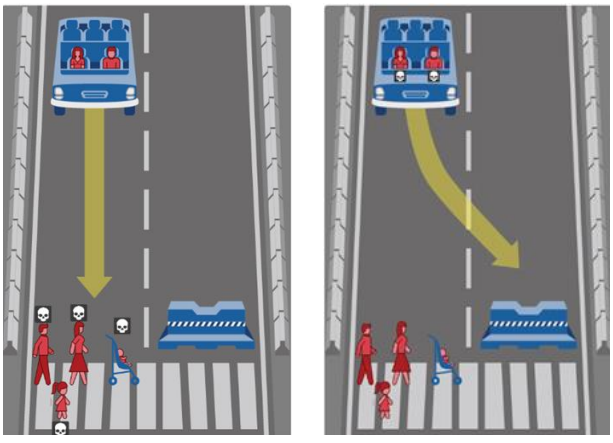
VII. Result and Analysis

A. Survey Responses

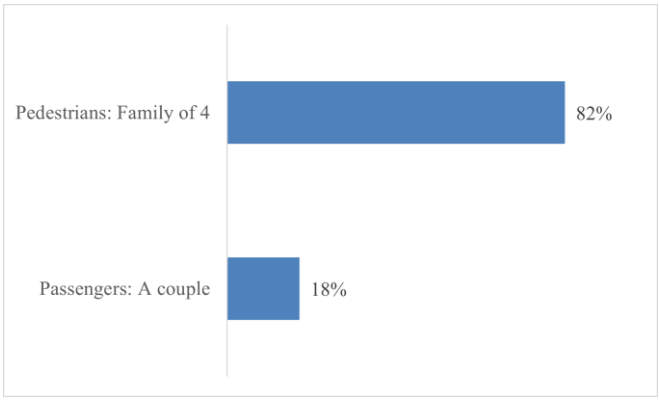
Question 1) Fully Autonomous Vehicles should be adopted and used on a day-to-day basis.

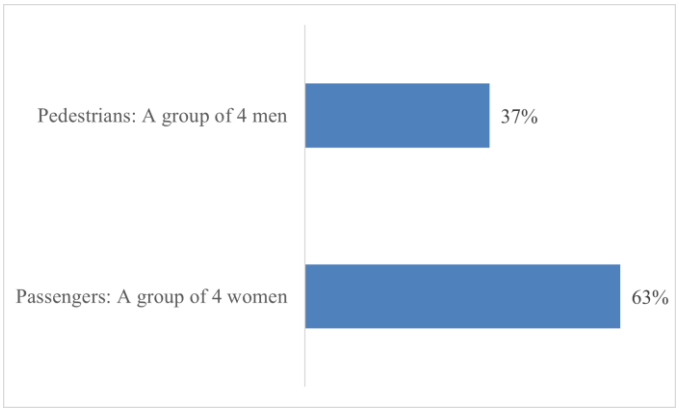


Question 2) Who would you rather save?

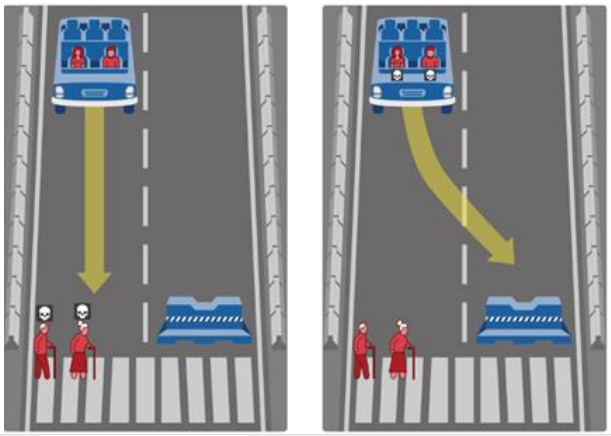


International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 08, August-2022

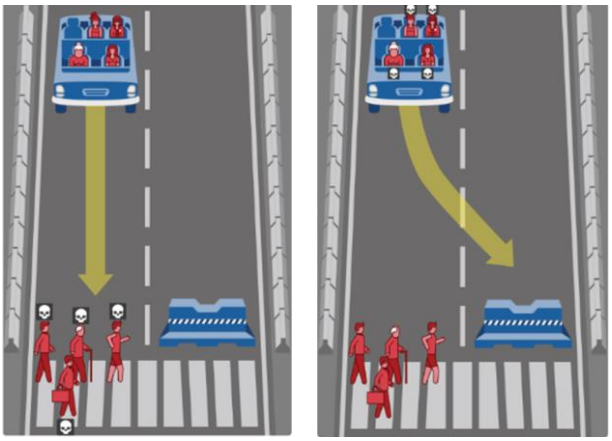




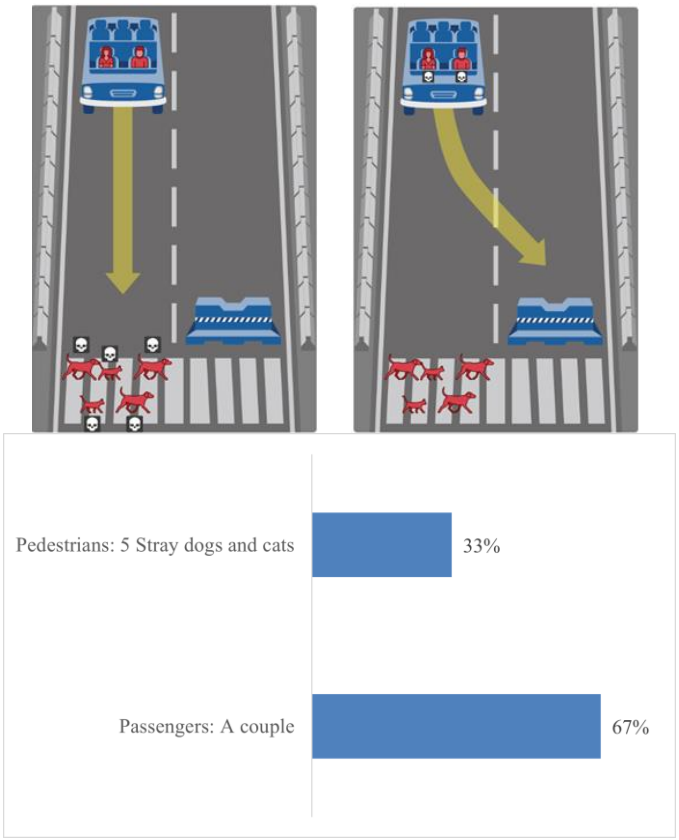
Question 3) Who would you rather save?



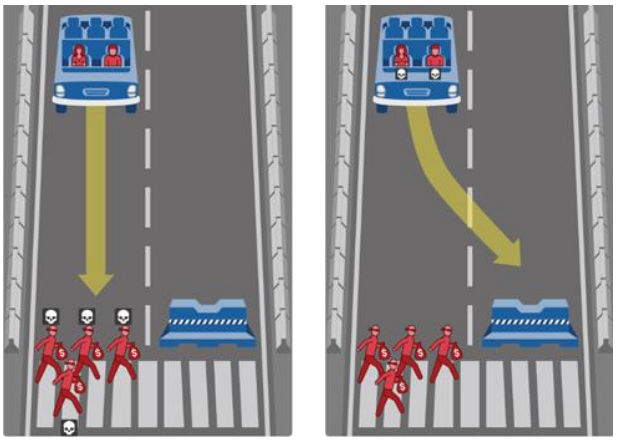
Question 4) Who would you rather save?



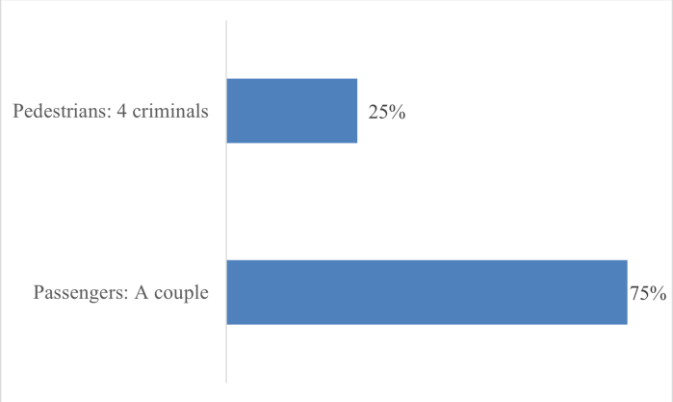
Question 5) Who would you rather save?



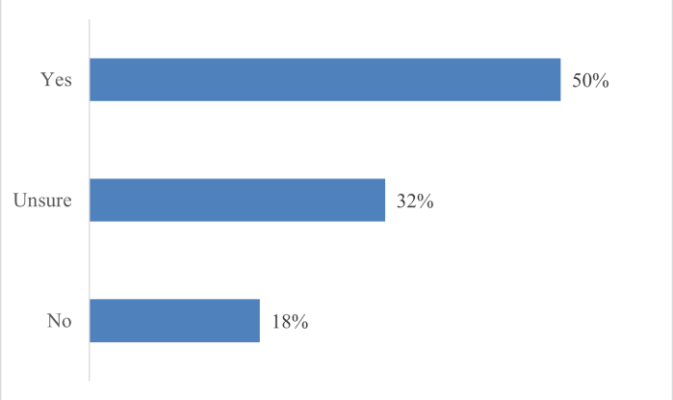
Question 6) Who would you rather save?



International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 08, August-2022



Question 7) Would any of your previous answers change if either of the pedestrians or passengers were your family members?



Question 8) Following up from the previous question, why do you feel that way?

This was the only subjective question from the survey and was only used for filtration. Those who answered yes briefly talked about how family is the most important thing to them and in such a case, they would choose to be selfish. Others questioned which family member was at risk while others stated that saving the most people is important – the ideology that should be followed by AVs to minimise damage to humanity.

B. *Possible Motivation for Answer Pattern*

Question 1: Most of the people were unsure whether they were in favour of Autonomous Vehicles or not. Generally, most people agreed that AVs should be adopted however almost 30% of the people disagreed. This may be due to the ethical considerations around AVs or the fact that they may take jobs away from humans.

Question 2: This was the first question that put human lives against each other. As expected, a significant majority of respondents chose to save more human lives, that is, the family of 4 people. However, some did choose the couple, perhaps because they imagined themselves in the passenger's situation and how they would be selfish and want to be saved, even at the cost of others.

Question 3: It was expected that the young couple would be saved as they have a longer life remaining; yet nearly 45% of the people chose the elderly couple. This might be due to the fact they feel sympathetic towards the elderly, and they feel the need to protect them.

Question 4: Out of all the questions, the outcome of this was the most difficult to predict but it had a clearer majority than the previous question, which was comparatively much easier to answer. Respondents chose to save the women, again because they might have felt sympathetic. It can also be due to the fact that generally women live longer than men, and hence, people felt the need to save them.

Question 5: In this case, instead of human lives against each other, a greater number of animal lives were put up against human lives. Despite more dogs and cats, most respondents chose to save the humans, possibly due to the fact that they would also want to be saved in this situation. They may feel more concern for humans than animals.

Question 6: In this situation, a greater number of criminals were against a couple. Most people saved the couple as they must not have felt that the criminals were worth saving, living ethically wrong lives. Although in real time, humans might not be able to identify criminals from regular people, in the future it is possible that Autonomous Vehicles, using facial recognition and extensive databases, can differentiate criminals and even important world leaders from other people.

Question 7: When family and other loved ones come into play, it is obvious that the opinions of people will change. Half of the people agreed their answers would change, showing how it is more efficient for an impartial AV to make these decisions for humans.

C. *Suggestions For Further Follow-Up Research to Be Done by Others*

In the survey presented, the respondents were all from the USA. To obtain results that truly reflect the entire globe's opinion and to have a healthy balance between "Eastern" and "Western" morals, it is necessary to conduct a bigger survey with respondents from all over the world. This survey only asked 6 situational questions whereas, many more variations can exist such that the majority answer might change. When testing for gender or socio-economic bias in whom to save, it is important to check if there are any changes in results if the passengers and pedestrians are interchanged; that is, in the case of question 6, the criminals become the passengers, and the couple the pedestrians. This might show that people generally favour the passengers no matter the people in the car or show the opposite. The survey should also be optimised to get insights into the motives of respondents for their answers.

VIII. CONCLUSION

As it is seen in the "trolley problem", when the conditions of the problem change, so does the "correct" answer. Utilitarianism aims at just saving the maximum number of people. However, there are other factors at play as well. As

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 08, August-2022

shown in the survey, it is imperative to consider the age, gender, number, and even the socio-economic factors related to both the pedestrians and passengers. Despite a somewhat clear majority in each of the questions, these questions will most likely stay unanswered and be debated. What convolutes the thinking when considering these situations is human emotion. Instead of looking at the problem with a purely objective view, varied factors, which ultimately should not be given more importance than saving the greatest number, lead to different answers. Thus, in such situations, human regret should be given a back seat and it is better for a completely objective party, such as an Autonomous Vehicle, to make such decisions for humans based purely on the number of people.

At the end of the day, it must be accepted that as the manufacturers of the autopilot software, the automotive industry will decide the answer. Considering such companies want to market their AVs as being the safest and most efficient choice, it is obvious they should program the software to save the passenger every single time, no matter the factors mentioned above.

IX. ACKNOWLEDGMENT

I would like to thank my advisor Mr. Gwyn Harold George Day.

X. REFERENCES

[1] *7 Benefits of Autonomous Cars*. (2017, July 21). Thales Group. https://www.thalesgroup.com/en/markets/digital-identity-and-security/iot/magazine/7-benefits-autonomous-cars

[2] Crockett, M. (2018, February 14). *The Trolley Problem: Would You Kill One Person to Save Many Others?* The Guardian. https://www.theguardian.com/science/head-quarters/2016/dec/12/the-trolley-problem-would-you-kill-one-person-to-save-many-others

[3] Kanter, Z. (2015, February 3). *How Uber's Autonomous Cars Will Destroy 10 Million Jobs and Reshape the Economy by 2025*. Zack Kanter. https://zackkanter.com/2015/01/23/how-ubers-autonomous-cars-will-destroy-10-million-jobs-by-2025/

[4] Kuipers, B. (2016, June 27). *We Need Ethical Robots. Asimov's Laws Are A Good Way To Start | GE News*. General Electric. https://www.ge.com/news/reports/beyond-asimov-how-to-plan-for-ethical-robots

[5] M. (2021, May 3). *Vehicle Dynamics: The Kinematic Bicycle Model*. The F1 Clan | An F1 Community Created by the Fans for the Fans. https://thef1clan.com/2020/09/21/vehicle-dynamics-the-kinematic-bicycle-model/

[6] Martinez-Diaz, M., & Soriguera, F. (2018). *Autonomous Vehicles: Theoretical and Practical Challenges*. Science Direct. https://www.sciencedirect.com/science/article/pii/S2352146518302606

[7] Othman, K. (2021, February 26). *Public Acceptance and Perception of Autonomous Vehicles: A Comprehensive Review*. Springer Link. Retrieved August 20, 2022, from https://link.springer.com/content/pdf/10.1007/s43681-021-00041-8.pdf

[8] Ribeiro, J. (2021, December 15). *Autonomous Vehicles : A New Mobility | Towards Data Science*. Medium. https://towardsdatascience.com/how-autonomous-vehicles-will-redefine-the-concept-of-mobility-582f8701a5f8

[9] Tikkanen, A. (2022, May 17). *Three Laws of Robotics | Definition, Isaac Asimov, & Facts*. Encyclopedia Britannica. https://www.britannica.com/topic/Three-Laws-of-Robotics

[10] Timmons, M. (2022, August 17). *Car Ownership Statistics in the U.S.* ValuePenguin. https://www.valuepenguin.com/auto-insurance/car-ownership-statistics