Ethics Analysis of Tesla's Autopilot Feature and Related Actions

STS Research Paper
Presented to the Faculty of the
School of Engineering and Applied Science
University of Virginia

By

Ishan Koroth

April 15, 2024

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.

ADVISOR

Benjamin J. Laugelli, Assistant Professor, Department of Engineering and Society

Introduction

In 2019, a fatal accident occurred when a Tesla Model 3 with Autopilot active had crashed into a truck at 70 mph, killing the driver, Jeremy Banner, on impact ("The final 11 seconds", 2023). The family of Banner had sued Tesla regarding the Autopilot feature having flaws, and the case at hand was to determine whether the driver or the actors behind the software were to blame. Several incidents involving crashes attributed to the Autopilot feature had been taken to the court. Determining the blame for these incidents has posed an ethical dilemma requiring analysis. Tesla won the case by arguing that human error, specifically the driver's lack of attentiveness while using Autopilot, was solely responsible. However, this decision remains a subject of ongoing debate.

Accidents involving autopilot have been occurring widely, with the NHSTA reviewing over 900 crashes in which Autopilot was used (Banker, 2023). Another aspect to consider is how the habits of people have changed as technology in autonomous vehicles have evolved, with people placing more trust on the vehicle to transport them to a particular location with less regard to the safety of their own and others that are affected by the potential risks. Even though many cases involving autopilot-related accidents put the blame on human error, it should also be considered that such cars should not have been rapidly produced to the masses had human error been a common causation factor in the first place. Standards placed by the NHSTA emphasize the need for full human engagement and control with the advanced features active, and despite that fact, accidents are still occurring due to numerous issues with the autopilot (NHTSA, 2024).

This case will be analyzed using the principles of duty ethics, specifically making use of categorical imperatives that guided the design of the Tesla Model 3's autopilot system. The universality principle will be used to evaluate whether the imperative could be considered a

universal law. This law is crucial for understanding the potential consequences that could affect the safety of anyone involved with or using the vehicles. The reciprocity principle will also assess the mutual relationship between Tesla and the consumers. To support my argument, I will analyze statements made by Tesla and Elon Musk, research done on the technology, accounts of personal experiences, and various cases involving accidents that involved the use of autopilot. By understanding Tesla's production and advertising practices, we can identify disconnects and pinpoint how the autopilot experience, ranging from sufficient to fatal, has occurred among users.

Literature Review

The following scholarly materials provide information regarding the decision-making and human factors behind a potential FSD system compared to that of the standard autopilot. This study was done to understand how people utilize vehicles depending on various levels of autonomy. Another scholarly article involves evaluating the human factors involved with regards to how people utilize Tesla vehicles while also diving into the good, bad, and ugly of various features that define the various strengths and weaknesses of Tesla vehicles. These two materials provide insights on the various factors that may have led up to points where disaster has occurred.

The first article focuses on the human factors that are negatively impacted by the increase in autonomy, with comparisons being done between an FSD beta system and the standard autopilot system, and the latter is currently being used in the Tesla Model 3 (Nordhoff et al., 2023). Experiments have been done on people of various backgrounds, and data was gathered by asking questions regarding their experiences with said vehicles. The results reflected upon how much work and awareness people had for both autopilot and FSD vehicles, with the latter

requiring a lot more attention than the former. This was due to the stress of having less control over the vehicle. For those who drove using the Autopilot, the drivers were more relaxed and felt a greater sense of convenience, which led them to not worry or care anymore as what mattered was reaching the destination. However, one driver noted that they had been alerted when their car drove over road debris, which forcefully reminded them to be an attentive driver rather than a passenger of the semi-autonomous vehicle.

The second article focuses more on the design aspect of Tesla and how the features affect the overall user experience in general. The main points that this article focuses on is how the auto-pilot functionality works along with other supplementary features (Gillmore & Tenhundfeld, N. L. 2021). One main aspect of the Tesla vehicle this article brings up is how the transfer of control (TOC) functions, and a major point is stated regarding the disengagement that is caused by the vehicle not being able to navigate properly in certain scenarios. As a result of disengagement, the vehicle will simply continue ahead as a normal car would. If the human is unable or unprepared to regain control, the risk of experiencing disastrous consequences to both the users and others largely increases. Something else that has been discussed was the need for a fail-safe in case the human is no longer engaged enough to regain control of the vehicle. This is because the driver, even if they may not be distracted with secondary tasks, may experience performance decrements, unfamiliarity, and close calls while driving with autopilot active. Such behavioral actions, which most humans in general adapt to, could attribute to risky behaviors and cause disasters.

These articles relate to one another such that complacency and over-reliance of the auto-pilot technology in the Tesla vehicles were highlighted as major human factors that negatively affected the driver. Another aspect to note is the transfer of control, and both articles touch upon

how the driver feels like a passenger since they do not feel the need to be as attentive with the vehicle taking over certain tasks. The feeling of relaxation and lack of care, combined with the potential risk of disengagement and improper faults in the autonomous functionality, have been factors attributed to the accidents that have occurred with the Tesla Model 3 when the autopilot was active.

Conceptual Framework

The conceptual framework that will be used in this case involves developing various categorical imperatives, derived from the duty ethics framework. A categorical imperative is a foundational principle of an action such that a person should be able to judge whether such action is morally right. This will be used to analyze the ethics behind decisions made in mass producing the vehicle and how they advertised/implemented the technology. The Kantian Theory, derived by Immanuel Kant, focuses on emphasizing actions that are morally right if it agrees to a moral rule or norm, which falls under the ideas of duty ethics and good will (van de Poel and Royakkers, 2011).

Immanuel Kant was an influential philosopher who initiated the way people thought of the world. Kant stated that it is not the fact that the world itself, but rather the human mind that provides the origin of experiences that people go through in their lives. He also proposed an idea such that morality is understood based on one's rationality, and two principles are derived based on this idea to evaluate the morality of a particular imperative or action.

One principle that will be investigated is the universality principle, which states that one should act on that maxim which would be fitting as a universal law that people would follow intuitively. Such a principle emphasizes the need to understand whether a particular action/decision is sensible and moral based on whether it benefits the people in a universal

manner. The action/decision must also fit the norm of good will perceived by the people collectively.

The second principle involves reciprocity, which involves treating humanity as a person and never as a means only. This principle is important in analyzing how the decisions behind the categorical imperatives were made with respect to the humans that experience the consequences because of the effects derived from the decisions made. Such consideration must be taken to evaluate whether the imperative was done to treat humanity well or simply to satisfy the objective at hand.

The following case will be analyzed with two categorical imperative principles in mind based on two actions made by Tesla. These will focus on whether it was right to mass produce the Tesla Model 3 with potential risks and defects. They will also be used to evaluate whether the advertisements of the autopilot technology were considered valid enough to allow people to make well-informed decisions. Such imperatives were selected to divide the factors in the decision making into specific categories to then be evaluated with the two principles derived from the duty ethics framework.

Analysis

With the principles and categorical imperatives derived from the conceptual framework, this case will be analyzed in depth to determine the validity of the actions made by Tesla. when they mass produced and advertised vehicles that posed a potential risk and have possibly misguided and misinformed consumers. The sections below will focus on first providing context based on primary sources that have been derived from statements and claims made by Tesla and other credible sources directly affiliated with Tesla. This will then be followed by a validity

analysis based on secondary sources using the universality and reciprocity principles for each categorical imperative.

Categorical Imperative #1: Mass Production of Vehicle with Potential Risks

Tesla's decision to mass-produce their vehicles with potentially faulty autopilot technology violates both principles of universality and reciprocity. Elon Musk had posted an article back in 2006 explaining his Master Plan in how he was going to eventually evolve Tesla into a company that mass produced affordable electric vehicles (Elon Musk, 2006). He started off by producing an electric sports car while promoting awareness for electric vehicles being pioneered for the masses. From there, Tesla went on to produce the Model S, X, 3, and Y to provide vehicles to people. In Elon Musk's intent to produce the vehicles, there is not much regard to the ethical implications involved when producing a vehicle to the masses. This brings upon a great concern as Elon Musk's intents focus mainly on the economic and environmental benefits at the expense of peoples' safety.

As time went on, the distribution of sales was heavily biased on the most affordable options, with the Model 3 being the best-selling model (Lambert, 2021). As a result of this, more people had access to the Model 3 and would make use of the vehicle for their transporting needs. From that point, people started to experience issues with the vehicle as numerous cases of accidents involving the Autopilot feature erupted, and the most notable case involved a Tesla Model 3 experiencing a fatal crash against a trailer of a truck ("The final 11 seconds", 2023). Since this accident and its thorough analysis on what happened has been publicized, this would mean that many people may have also read upon the case and recognized the risk in owning a Tesla vehicle. To elaborate on accident rate statistics, a study conducted by Lending Tree ranks Tesla as the number 1 vehicle for highest accident rates compared to various other brands, and

this shows how much of an impact the vehicles and technology have on people in general (Ranz Injury Law, 2024). With cases of accidents involving Tesla's autopilot rising, this would affect several of those who own the vehicles as they slowly lose trust in the autopilot technology. Along with the fact that more people have been owning Tesla vehicles over time, the cases and statistics would only grow more apparent and exacerbate mistrust and fatalities.

Universality Principle Validation

The imperative of mass producing the vehicle that possessed a fault violates the universality principle. This imperative can be evaluated with the said principle as the mass production of Tesla vehicles, especially the Model 3, allowed more people to access the vehicle and its technology. This fact derives a well-amounted sample that can be analyzed. Based on some verdicts of the numerous cases, the court declared that the accidents occurred due to human error and pointed out the fact that humans had to be engaged and attentive to the road while the vehicle is in an operable state (Hawkins, 2023). In one such case that declared Tesla as not liable, Micah Lee was killed after he crashed into a tree while driving his Tesla Model 3. The vehicle somehow veered off the road while having autopilot active.

In a worst scenario such as this one, the autopilot feature requires the human driver to be fully attentive and responsive. However, if the technology requires full attention, then that defeats the purpose of using autopilot. Even though Tesla uses the maxim of the driver needing to be responsible as a way of justifying their stance of the universality principle, companies in general should be able to deploy products with faults and place the blame on the users, even if the products themselves cause problems and potential fatalities. Thus, the mass production of the Tesla Model 3 possessing a potential risk violates the universality principle as it had proven to be commonly involved in multiple accidents and casualties. Had this been the general norm for all

people, the likelihood of accidents would only increase from this point onward. Such a trend would not be acceptable by the universality principle since it would generally be agreed that an increase in accidents due to the following is not a desirable norm.

As stated in this validation, the mass production of a vehicle that may have a fault is considered a violation of the universality principle since this action exploits the principle itself to put the blame on others at the expense of peoples' safety, which contradicts the general norm of universal safety. Certain court decisions, however, claim that it is not the fault of technology but rather a human error that is considered the root cause of the problem since the drivers are responsible for their actions on the road. To counter this alternate viewpoint, a court case was settled by Tesla regarding an accident when a potential fault was found with the vehicle veering off the road while using Autopilot (Zachariah, 2024). This evidence proves that human error is not the sole reason behind the accidents.

Reciprocity Principle Validation

Such an imperative violates the reciprocity principle as many people are paying money to purchase Tesla's vehicles for the numerous technological features offered, such as autopilot. However, they are not properly being reciprocated with the technology they sought for. With the Tesla Model 3 being the most common vehicle, more people would have access to such a vehicle. In a comment on a blog post discussing a fatal accident between a Tesla Model 3 and a trailer of a truck, as user had complained about the quality of the vehicles along with the TOC and Autopilot features lacking trust between human and vehicle (llama-lime, 2019). One notable criticism involved the vehicle experiencing issues in properly detecting obstacles of lighter color. In the case of the truck accident, the trailer of the truck was white. Other technological concerns included the lane-keeping technology experiencing issues detecting lane-markings, and even if

the driver had their hands on the wheel while the feature was active, the driver would still be notified to apply more resistance to the steering wheel.

This resulting dissatisfaction when using the vehicles violates the reciprocity principle as Tesla mass produced a vehicle with potential risks for the purpose of selling vehicles to as many as possible with less regard to the risks in the technology that was being deployed. By taking this action, Tesla had focused more on the means to gain profits and did not carefully consider the ends that people desired from the autopilot technology. With a sense of mistrust towards Tesla and the autopilot growing, certain people will no longer feel that their needs have been properly reciprocated by the mass deployment of Tesla's vehicles.

Categorical Imperative #2: Advertisement of Technology not being Informative

Tesla's decision to advertise their autopilot technology in an improper manner violates both principles of universality and reciprocity. The autopilot feature had been advertised to be sufficient in providing an overall proper driving assistance system. According to the autopilot subpage on the Tesla website, the autopilot technology has a high coverage with its sensors and possesses great processing power to gather data when autopilot is active (Tesla, 2024a). Another autopilot-related article goes into detail with regards to the various capabilities that the Tesla autopilot possesses, such as traffic-aware cruise control, auto-steering, traffic and stop sign control, and other features that ease the responsibilities of the driver (Tesla, 2024b).

With regards to statements made, Elon Musk made claims regarding the Tesla's autopilot technology as he states that it is "probably better than a person right now" since he already believes that the technology has proved itself (Frankel, 2016). This statement reflects upon the bold nature of Tesla while also revealing a slight implication, especially with the word "probably" being used in the quote. Such a use of a word would mean that there must be

something that the people may not be fully informed about the autonomous capabilities presented. Elon Musk also claims that the technology will go through rapid development in the next couple of years. Considering the recent events involving the use of Autopilot, it seems to be the case that the misleading advertisements for the Tesla vehicles could have led to misinformed decisions.

Universality Principle Validation

The advertisements that provided misinformed statements regarding Tesla's autopilot violates the universality principle. Since the advertisements were spread out to the public, the universality principle can be applied to analyze the large sample of people who purchased the vehicles. A specific case that ended in Tesla having to provide settlement to the prosecutors for a fatal accident will be used an example of a part of that sample. The case being presented here involved an Apple engineer, Walter Huang. driving a Tesla Model X that would steer into a safety barrier, killing the engineer (Zachariah, 2024). The family of Huang would take the case to the court and claim that the autopilot feature was to blame for the crash. Huang's family had even questioned Elon Musk's claim as they felt it exaggerated the capabilities of Tesla's autopilot feature. Even though Tesla stated that the driver had his hands off the wheel and was distracted, it was also stated that the vehicle's autopilot failed to keep the vehicle in the lane. This highlighted a potential fault in Tesla's autopilot technology.

As a result of this, Tesla reached a settlement with this case, which was not the case in earlier fatal accidents. This case proves that Tesla had violated the universality principle as Tesla had promoted the autopilot with the expectation that people would pay more attention to the road. When looking into the multitude of recent cases that involve autopilot-related accidents, such a norm would not be sustainable and agreed upon in the real world due to safety concerns.

Another point to consider is the fact that previous cases involving fatal autopilot-related accidents had not reached a settlement due to evidence indicating that the driver was not paying attention. With Huang's case being settled, the autopilot feature in the previous cases may have malfunctioned similarly such that the vehicle could have failed to slow down in time and not remain in the lane. The insights delved from the case contradict the statements made regarding how the autopilot feature is supposedly better than a human, and in this case, the autopilot was no different.

Reciprocity Principle Validation

The advertisements regarding the Tesla vehicles and technology that offer people features that they may desire from the Autopilot for convenience violate the reciprocity principle. Reciprocity applies in this case such that the people must both be satisfied and well-informed with the decision made in purchasing the vehicle. To state a specific case related to reciprocity, Angela Chao, CEO of a Chinese shipping firm, had drowned into a pond while reversing in a Tesla Model X (Hernandez, 2024). This tragic event was attributed to human error as Chao had been intoxicated while driving the vehicle. From such a point alone, Tesla could have argued that the driver was at fault since the driver lacked the attentiveness to drive the vehicle.

Even if human error seemed to have been the likely case here, one thing to note is why Chao made the decision to drive back in the Tesla. If the vehicle did not possess autopilot technology, perhaps Chao would have made an alternate decision for the sake of her safety. Considering the claims made by Elon Musk regarding the autopilot technology being better than that of a human, it can also be inferred that Chao had been falsely misguided by the statements. She may have felt that her Tesla vehicle was the better option over a fellow friend.

This case represents an example of a reciprocity principle violation as Chao had her life compromised rather than being treated with proper care by Tesla's efforts to promote their autopilot technology to the masses. Despite implementing features to the autopilot that offer convenience to the driver, cases of accidents have still been on the rise that demonstrate violations of the reciprocity principle in failing to deliver what had been advertised. Had more emphasis been placed on accomplishing the ends rather than the means in deploying the autopilot technology, the lives of many, including Chao, could have been saved.

Conclusion

The autopilot technology implemented by Tesla has been the center of various debates regarding ethical and moral concerns for peoples' safety. The reviews, accidents, and various other cases presented showcase examples of Tesla failing to mass produce a vehicle that delivered their claims and advertisements advocating for their vehicles and autopilot technology. Even though the Tesla Model 3 was not the vehicle that was involved in every case, the same can be said for the higher-end models that possess the same auto-pilot technology. The duty ethics conceptual framework, which divides certain imperatives behind the deployment and validates the universality and reciprocity of each, proves that Tesla made unethical decisions when handling the cases. This is also considering how the various cases were handled with less regard to potential faults on their end. Understanding the impact of how vehicles with autonomous capabilities affect the environment on the road is important since the safety of the driver and people on the road are put at risk.

Word Count: 3750

References

Banker, S. (2024, February 20). Tesla Has The Highest Accident Rate Of Any Auto Brand. *Forbes.* https://www.forbes.com/sites/stevebanker/2023/12/18/tesla-has-the-highest-accident-rate-of-any-auto-brand/

Frankel, T. C. (2016, January 11). Elon Musk says Tesla's autopilot is already 'probably' better than human drivers. *The Washington Post.* https://www.washingtonpost.com/news/the-switch/wp/2016/01/11/elon-musk-says-teslas-autopilot-is-already-probably-better-than-human-drivers/

Gillmore, S. C., & Tenhundfeld, N. L. (2021, February 9). The Good, The Bad, and The Ugly: Evaluating Tesla's Human Factors in the Wild West of Self-Driving Cars. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 64(1). https://doi.org/10.1177/1071181320641020.

Hawkins, A. J. (2023, October 31). Tesla wins another court case by arguing fatal Autopilot crash was caused by human error. *The Verge.* https://www.theverge.com/2023/10/31/23940693/tesla-jury-autopilot-win-liable-micah-lee

Hernandez, E. (2024, March 21). Foremost CEO Angela Chao drowned after Tesla sank into pond. *My San Antonio.* https://www.mysanantonio.com/news/local/article/angela-chao-cause-of-death-18967658.php

Lambert, F. (2021, February 4). Tesla Model 3 is still the best-selling electric car in the world and it's not even close. *Electrek.* https://electrek.co/2021/02/04/tesla-model-3-still-best-selling-electric-car-world-not-even-close/

llama-lime (2019, May 16). I recently got a Model 3, and while I love it, and love the adaptive cruise control, I don't trust [Comment on the blog post "Autopilot was active when a Tesla crashed into a truck, killing driver"]. *Ars Technica.* https://arstechnica.com/cars/2019/05/feds-autopilot-was-active-during-deadly-march-tesla-crash/?comments=1&comments-page=5#comment-37374595

Musk, E. (2006, August 2). *The Secret Tesla Motors Master Plan (just between you and me)*. Tesla. https://www.tesla.com/blog/secret-tesla-motors-master-plan-just-between-you-and-me

NHSTA. (2024). *Automated Vehicles for Safety*. https://www.nhtsa.gov/vehicle-safety/automated-vehicles-safety

Nordhoff, S., Lee, J. D., Calvert, S. C., Berge, S., Hagenzieker, M., & Happee, R. (2023, February 23). (Mis-)use of standard Autopilot and Full Self-Driving (FSD) Beta: Results from interviews with users of Tesla's FSD Beta. *Frontiers in Psychology, 14(Sec. Human-Media Interaction).* https://doi.org/10.3389/fpsyg.2023.1101520

Ranz Injury Law. (2024, March 11). *Why are Tesla car accident rates so high?.* https://www.ranzlaw.com/why-are-tesla-car-accident-rates-so-high/

Tesla. (2024a). *Autopilot*. https://www.tesla.com/autopilot

Tesla. (2024b). *Autopilot and Full Self-Driving Capability.* https://www.tesla.com/support/autopilot

Thadani, T., Lerman, R., Piper, I., Siddiqui, F., & Uraizee, I. (2023, October 6). The final 11 seconds of a fatal Tesla Autopilot crash. *The Washington Post.* https://www.washingtonpost.com/technology/interactive/2023/tesla-autopilot-crash-analysis/

van de Poel, I., & Royakkers, L. (2011). *Ethics, Technology, and Engineering: An Introduction (1st ed.)*. Wiley-Blackwell.

Zachariah, B. (2024, April 11). Fatal Tesla "Autopilot" crash of Apple engineer reaches settlement – report. *Drive.* https://www.drive.com.au/news/fatal-tesla-autopilot-crash-reaches-settlement/