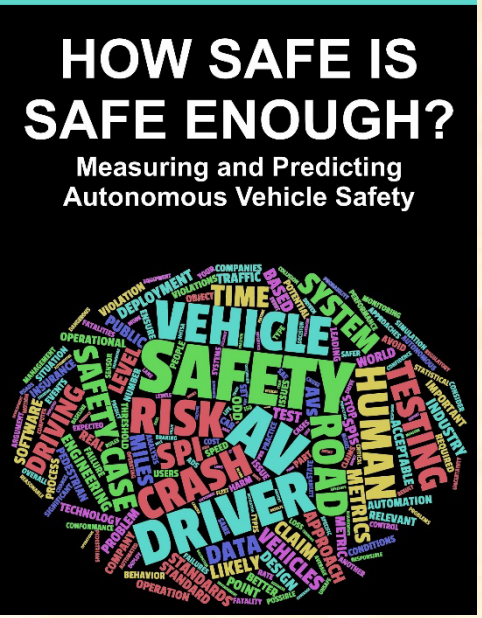
PHILIP KOOPMAN



How Safe Is Safe Enough for Autonomous Vehicles?

Prof. Philip Koopman

March 31, 2023

Full-length keynote version:
Nov 13, 2022 at https://safeautonomy.blogspot.com/

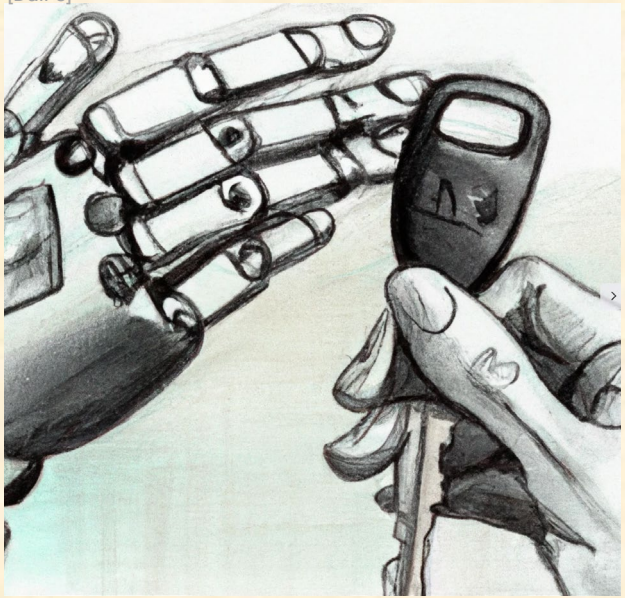
Carnegie Mellon University

Carnegie Mellon University

Overview

- Risk management frameworks
  - Which human is a baseline driver?
  - Risk mitigation is not safety
- A broader view of Safe Enough
  - Ethical considerations
  - Hierarchical model of safety needs
- Deployment criteria

[Dall-e]



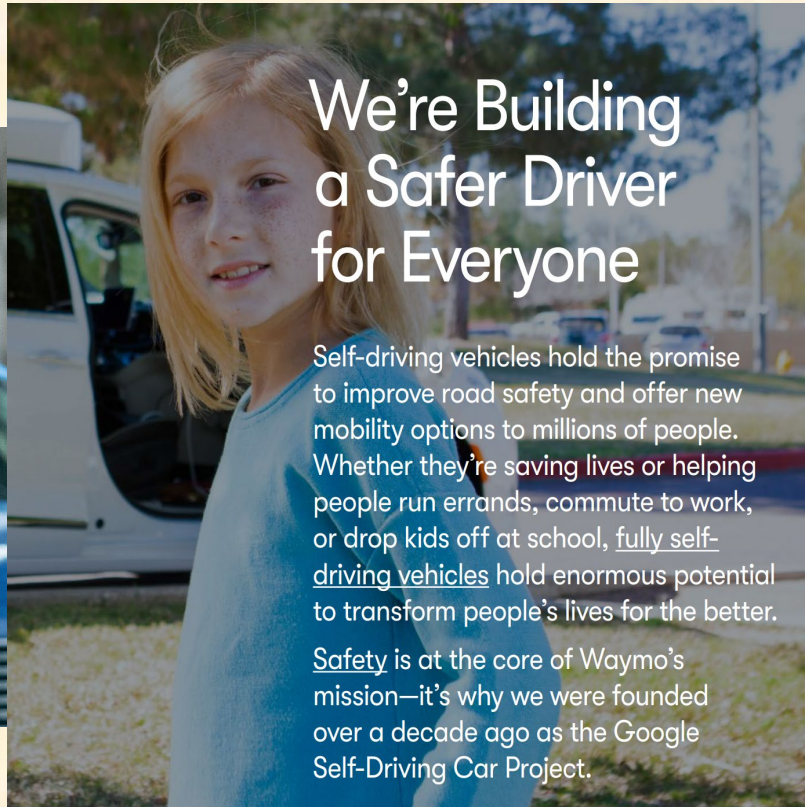
ADS = Automated Driving System

(Car drives; people can sleep)

ADS Technology: Sold Based on Safety

Ford VSSA https://bit.ly/3njionT

Waymo VSSA https://bit.ly/2QuYhai



Setting The Risk Goal

- MEM – Minimum Endogenous Mortality
  - System risk has minimal effect on overall risk
- ALARP – As Low As Reasonably Practicable
  - Reduce identified risks unless cost is extreme
- NMAU – "Nicht Mehr Als Unvermeidbar"
  - Reduce identified risks within reasonable cost
- SIL – Safety Integrity Level approaches
  - Engineering rigor applied to mitigate risks

[Dall-e]

- GAMAB – "Globalement Au Moins Aussi Bon"
  - At least as good as an existing system (e.g., a human driver)

Positive Risk Balance (PRB)

- Utilitarian GAMAB approach
  - 36,096 fatalities (1.10/100M miles)
  - 2,740,000 injuries
  - 6,756,000 police-reported crashes
  - Data includes drunk drivers, speeders, no seat belts

[DOT HS 813 060 & DOT HS 813 021]

2019 Data

➔ Expect zero deaths in a 10M mile testing campaign

- The averages do not necessarily apply
  - Which driver?
  - Under what conditions?
  - Driving which vehicle?

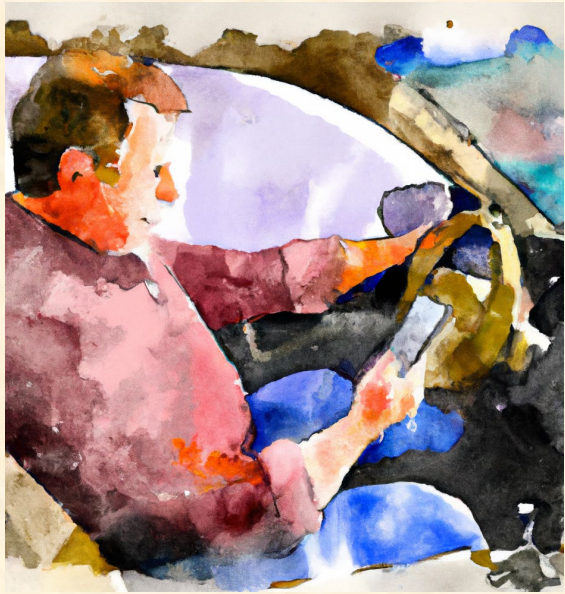
Which Driver Are We Better Than?

- ~100M miles/fatal mishap for human drivers
  - 28% Alcohol impaired/Driving Under Influence
  - 26% Speed-related
  - 9% distracted driving
  - 2% drowsy …

[DOT HS 813 060 & DOT HS 813 021]

(total > 100% due to multiple factors in some mishaps)

- Fully functional drivers are much safer
- New AV has better safety than 10+ year old "average" car

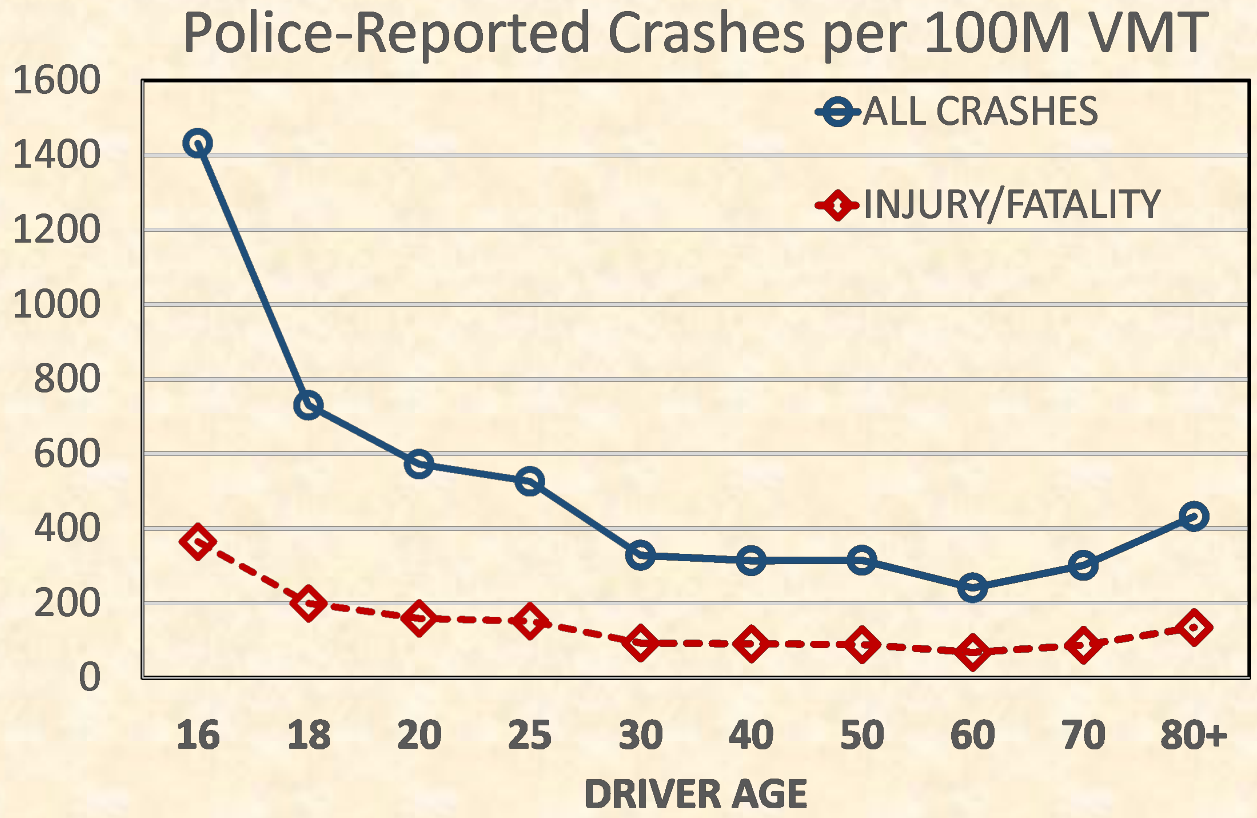


[Dall-e]

➔ Better than an unimpaired, undistracted driver in new car

Driver Age Affects Crash Rates

➔ Better than a middle-aged driver



Region Affects "Safe Enough" Value

- Fatality averages for 2019 (IIHS)

Location	Deaths/100K people		Deaths/100M miles		
DC	3.3	} 7.7x	MA	0.51	} 3.4x
US	11.0		US	1.11	
WY	25.4		SC	1.73	

- Fatal crash type [IIHS Fatality Fact Sheets State by State; DOT HS 813 060]
  - DC: highest pedestrian rate (39%)
  - NY, FL, DE: highest bicycle rate (5%)
  - Fatalities per 100M miles: Urban 0.86 vs. Rural 1.65
  - What about day/night, weather, etc.?

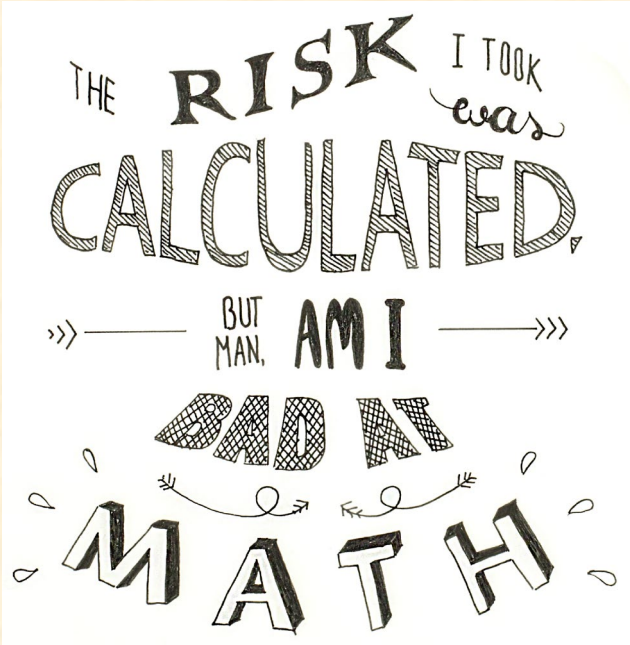
https://bit.ly/3CJm7nP

➔ Better in same conditions as AV operations

Ethics: Deployment Governance

- #1 ethical issue in AVs is deployment governance
  - Who decides when to deploy based on what?
  - Missing independent technical oversight
  - Huge pressure for aggressive deployments

- Safe enough to deploy should address:
  - Publicly disclosed safety prediction
  - Inclusion of stakeholder concerns
  - Transparency of data & processes
  - Accountability for any losses
  - Non-discrimination in operational concept



https://bit.ly/3rJeaJ4

What People Mean By “Safe”

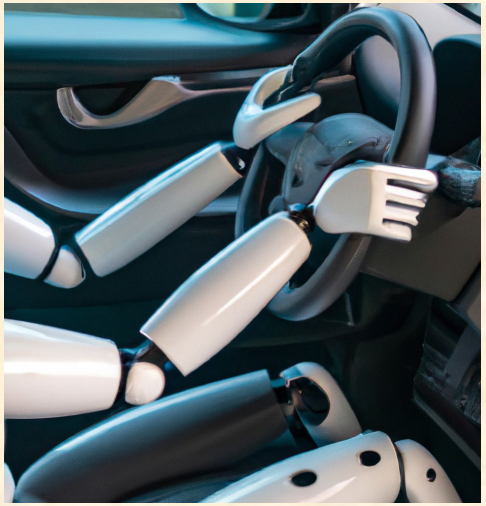
- Human drivers are bad, so computers will be safe
  - Industry rhetorical talking points are ubiquitous
- “Safety is our #1 priority”
- Safe driving behavior
  - Follows traffic laws; good roadmanship
- Tested/simulated for millions of miles
- Risk is managed via insurance
- Conforms to safety standards
- Positive Risk Balance
- Safety cases supported by evidence

[Dall-e]

Hierarchy of Concurrent Safety Needs

Elements of Safe Enough AV Deployment

- Deployment Governance
  - Stakeholders involved in criteria & decision
  - Safety culture assures fair dealing on decision
- Acceptable risk
  - Safe expectations + extra margin for unknowns
  - Safety & security industry engineering standards
  - Ethical concerns addressed
- Safety case
  - Transparent argument based on evidence
  - Lifecycle uncertainty management via field feedback
- Safety while public road testing



[Dall-e]

Key Observations

- Safe Enough is not a single number
  - Compared to which driver, when, where?
  - “Safer than human” is just a starting point
- Different aspects of safety to address
  - Ethical considerations
  - Management of uncertainty
  - Lifecycle safety support
  - Should safety standard conformance remain optional?
- Deployment governance is #1 ethical issue
  - Who decides “safe enough” based on what data?

https://on.gei.co/2r2rjzg

How safe is safe enough for Autonomous Vehicles

by Philip Koopman, 03/31/23

Yen-Chu Wu

Part I. Literature

Koopman, P., & Wagner, M. (2017). Autonomous vehicle safety: An interdisciplinary challenge. *IEEE Intelligent Transportation Systems Magazine*, *9*(1), 90-96.

Koopman, P., & Wagner, M. (2016). Challenges in autonomous vehicle testing and validation. *SAE International Journal of Transportation Safety*, *4*(1), 15-24.

Liu, P., Yang, R., & Xu, Z. (2019). How safe is safe enough for self‐driving vehicles?. *Risk analysis*, 39(2), 315-325.

Rodionova, A., Alvarez, I., Elli, M. S., Oboril, F., Quast, J., & Mangharam, R. (2020, October). How safe is safe enough? Automatic safety constraints boundary estimation for decision-making in automated vehicles. In *2020 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1457-1464). IEEE.

Shariff, A., Bonnefon, J. F., & Rahwan, I. (2021). How safe is safe enough? Psychological mechanisms underlying extreme safety demands for self-driving cars. *Transportation research part C: emerging technologies*, *126*, 103069.

Part II. Recent News

DaSilva, S. (2023, March 4). Human Drivers Avoid Crashes 99.999819% of the Time, Self-Driving Cars Need to Be Even Safer. *Jalopnik*. https://jalopnik.com/self-driving-car-vs-human-99-percent-safe-crash-data-1850170268

Rasa G, (2023, March 2). AAA study says Americans have grown more afraid of self-driving cars. Good! *Autoblog*. https://www.autoblog.com/2023/03/02/aaa-autonomy-study-missy-cummings-lecture-university-of-michigan/

Vallance, B. C. (2022, August 18). Driverless cars: Experts warn no easy answer to how safe they should be. *BBC News*. https://www.bbc.com/news/technology-62598618

Part III.

Jinhua questions

Q: People and machines make mistakes in different styles. How does that mean for us to regulate or set up a standard for autonomous vehicles versus conventional vehicles? Why have a standard for autonomous vehicles as opposed to everything else?

A: There are standards for conventional vehicles, but even the car industry doesn't publicly commit to following them. He also points out that the safety standards for autonomous vehicles are tricky because traditional testing methods don't apply when using machine learning. Instead of outlining a specific process or method, the standards focus on telling a story to demonstrate understanding and not missing any important details. It's a different approach than traditional safety standards.

Q: Safety is about the bad days instead of the average case. Whether there should be different criteria for AV acceptance in different locations, similar to the differences in emissions standards between California and other states. Whether there should be a consistent standard across the board or if it is appropriate to have varying standards in different locations?

A: There is no fixed criterion. It depends on whether the car company can define what "safe enough" means in a way that makes sense and if they can achieve a standard of X times safer than humans. However, if they want to cut it closer, they would have to invest time and energy justifying why they're cutting it so close. Even if the aggregate number is good, the distribution of harm to demographics matters, and redistribution may be necessary. The speaker gave an extreme case to emphasize the point.

Q: There are two extreme examples: (1) products cannot be sold until they are proven to be valid, like with the FDA and medicines, and (2) software can be deployed and problems can be fixed later with patches. Where autonomous vehicles should fall on this spectrum and how to reconcile the two extremes?

A: The risk involved in autonomous vehicle testing affects other road users who did not sign up for it. It is not justifiable to increase the risk to other road users for the sake of getting autonomous vehicles on the road faster.

Audience's question

Q: Which internationally influential regulator will define the industry in the end? Because in the end, the car industry wants to sell cars everywhere. Which regulator will win will define the standards and formulate the final attitudes?

A: It is unclear which regulator will define the standards for the industry, as regulators can set whatever standards they want, and if there are big crashes that kill important people, those standards will change. The history of safety agencies shows that they happen in response to big, high-profile media loss events, and there is no reason to believe that automated vehicles will be any different. Therefore, the presenter does not know who will win in the end.

Q: What type of evidence would you want to have from the car industry before they are truly allowed to deploy at scale?

A: The US NHTSA proposed a rulemaking in December 2020 that requires following industry standards and presenting a safety case that proves the vehicle is safe enough for deployment. However, this proposal has been stagnant since then.

Q: You've worked with both industries on the culture of safety in aviation versus the culture of safety in both the traditional automotive industry and this emerging autonomous industry. Do you have any big-picture reflections or observations on the differences in how they approach safety regarding the regulation?

A: The automotive industry lacks a strong safety culture and is hesitant to follow safety standards, unlike in aviation where safety is a priority. The safety standards don't get talked about much by car industries. Even if they do, they are afraid to make it a requirement for doing business.

Q: People believe AV will need to be significantly safer than the average human driver to gain adoption. Many people view errors by humans as inherently more acceptable than those by computers. Do you agree?

A: Autonomous vehicles need to be 100 times better than human drivers, with two factors of 10. The first factor is for uncertainty in performance and the second is for the redistribution of risk. There are many unknowns and unknown failures that could have real-world consequences, which may require another factor of 10. A factor of 100 is necessary to have a decent confidence interval.

Q: If all vehicles were autonomous, can we achieve that 100 x order of magnitude increase in safety that you just mentioned?

A: If it's completely automated, that helps reduce the incidence of human error. It will take decades before all conventional vehicles are retired, and even then, there will still be pedestrians, light mobility users, and other unpredictable factors on the road. Another point is that humans are capable of compensating for other people's errors, which is a skill that goes beyond simply following the rules of the road.

Q: Tell us more about the concept of deploying vehicle-to-vehicle-to-infrastructure communications to enhance overall system and safety?

A: While external information can improve safety, the baseline system must be entirely self-sufficient and acceptably safe without it. External information may not always be available, it may be broken or unreliable, and infrastructure will never be pervasive. Therefore, the system cannot rely solely on external information and must be able to operate safely on its own.

Part IV. Summary of Memos.

Themes from Other Memos

1. Nineveh reflected on Philip Koopman's talk on autonomous vehicles and questioned the purpose of creating them if the goal is not to make road transportation safer. During the discussion, other potential benefits of AVs such as meeting latent demand for senior citizens, youth, and other demographics, freeing up road space, and making transportation more affordable and accessible were discussed, but all come with potential negative externalities or could be feasibly addressed without automation. Lastly, Nineveh suggests exploring other alternatives like changes in road design and stricter driver license policies to achieve the goal of increased safety.
2. Michael found Koopman's talk on the safety of autonomous vehicles thought-provoking, particularly the idea of what it means to be "as safe as human drivers" and the need for AVs to be tested many collective miles before a comparable safety level can be reached. He also highlighted concerns about the deployment of AVs, including potential segregation of the urban environment based on AV zones and non-AV zones, and the different liability structures between human and machine mistakes. Michael believes that AVs should help humans without becoming "smarter" than them and that the broader discussion of the human-AI relationship is a pressing topic for society to resolve.
3. Yunhan believed that safety agencies should proactively establish rules and guidelines to ensure the safe development of autonomous vehicles rather than enacting regulations in response to incidents after they occur. Yunhan also discovered that regulatory practices associated with autonomous vehicles can significantly vary among nations and suggests establishing a "case library" to aid in creating a standard set of best practices and guidelines for the development and deployment of autonomous vehicles, regardless of where they are being tested.
4. Spencer found Professor Koopman's talk on determining the safety standard for autonomous vehicles interesting, particularly how human perception of safety is disentangled from reality. He pointed out that regardless of actual safety statistics, the more control an individual has over a particular mode of transportation, the more likely people are to perceive it as safer. Addressing this perception gap is crucial for the successful integration of autonomous vehicles into our transportation systems.
5. Jason reflected on Phil Koopman's discussion on how different risk management frameworks can be applied to this task. While some frameworks like minimum endogenous mortality may not work for autonomous vehicles, Koopman discussed others like ALARP and SIL. However, Jason notes that in the case of new technologies, like autonomous vehicles, quantitative evidence may not be enough to convince skeptics, and trust is an important factor in adoption. Jason suggests that reducing risks is still a crucial goal, but that it may not necessarily correlate closely with building trust among the public.
6. Ao reflected on the safety and ethical dimensions of autonomous vehicles (AVs). Koopman stressed the need for AV companies to build a safer driver and establish trust in the technology. Ao found it interesting to consider the contrast between a relative measure of AV safety for individual purposes versus an absolute measure for societal impact. The challenge of reconciling varying perspectives at the individual and collective levels remains a grand challenge.
7. James questioned whether society will accept occasional mishaps even if AVs achieve a high level of safety on a macro scale. He also raised the issue of individual rights in a future where AVs are substantially safer than traditional vehicles and governance structures attempt to stop the use of non-AVs, and whether individuals should have the right to drive themselves in such a future.

8. McKenzie found Phillip Koopman's discussion on the distribution of safety incidents and how it will change under AVs to be the most interesting point. However, she thought Koopman missed an important point when discussing the international regulatory agency that will determine safety standards, which may vary by country. Additionally, McKenzie was curious about how road design will change with the deployment of AVs and the perception of safety. She wondered how safety standards for roads will evolve to change the safety standards and performance of autonomous vehicles.
9. Jay was shocked to learn that there is no technical oversight for AV deployment in the United States and insufficient technical guidance in Europe. Koopman drew an important distinction between safety engineering and system engineering, with the core principle being "identifying hazards as a primary first-class activity." Jay believed that there needs to be stringent government oversight over the AV industry.

My Reflection

When it comes to autonomous vehicles, I always think and am concerned about safety issues. As humans, we make mistakes, but it feels like we have the power to control things when driving by ourselves. On the other hand, entrusting traffic safety and lives to machines or technology can bring about a certain degree of unease. I have been wondering about how much safety is considered safe enough as the title of this week's talk. Is the current technology capable of handling various emergent, unpredictable, and rare situations? Human-driven cars have a certain death rate, and we expect autonomous vehicles to be safer. However, whether a near-zero-death rate is realistic or not?

Even if technology advances, I still have some doubts about the ability of autonomous vehicles to respond to various emergencies. While computers do not suffer from issues like drunk driving, fatigue, or lack of concentration, I wonder how reliable they are in practice. Like using electronic products in daily life, is there a possibility for an autonomous vehicle to suddenly crash? My other concern is the responsibility for accidents involving autonomous vehicles. If an accident were to occur, who would be held accountable? Would it be the vehicle manufacturer, the software developer, or the owner of the vehicle? In cases where the vehicle's systems are hacked, who would be responsible for ensuring the safety of the passengers and other road users? I believe these concerns should be addressed before fully embracing autonomous vehicles.

Overall, the presentation provided me with a lot of insights and sparked my interest in thinking more about the safety and ethics of autonomous vehicles.