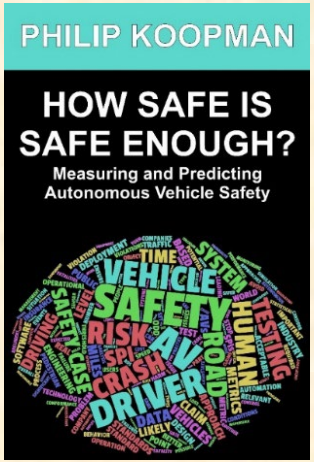
Prof. Philip Koopman

Truths & Myths About Automated Vehicle Safety



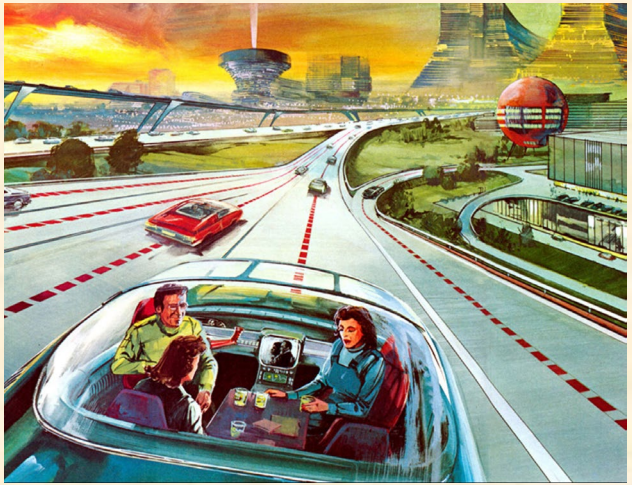
Carnegie Mellon University

www.Koopman.us

Overview: Automated Vehicle Safety

Carnegie Mellon University

- Sorting out truth, myths, and "it's complicated"
  - Companies say they are safer than human drivers
  - But public trust has been eroding
- Truth/Myth topic areas, including:
  - Are automated steering features safer?
  - Are robotaxis safer than humans yet?
    - Is that even the right question to be asking?
  - Important misconceptions
  - Other issues that still need attention



[General Motors]

© 2024 Philip Koopman

Why Is AV Safety Complicated?

Carnegie Mellon University

- Public expectations
  - Expect super-human machine performance
  - Trust too easily given, backlash when broken
- Technical challenges
  - Machine Learning safety is work in progress
  - Statistical approach vs. high severity rare events
- Industry culture clash
  - Machine Learning: 99% is a great result vs. safety is 99.9999…%
  - Silicon Valley: move fast + break things
  - Automotive: blame driver for not mitigating equipment failures
  - Regulators: test-centric; struggling with software safety

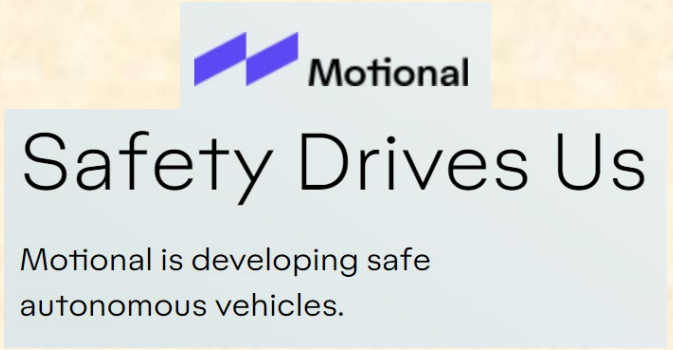
© 2024 Philip Koopman

Robotaxis: “Safety Is Our #1 Priority”

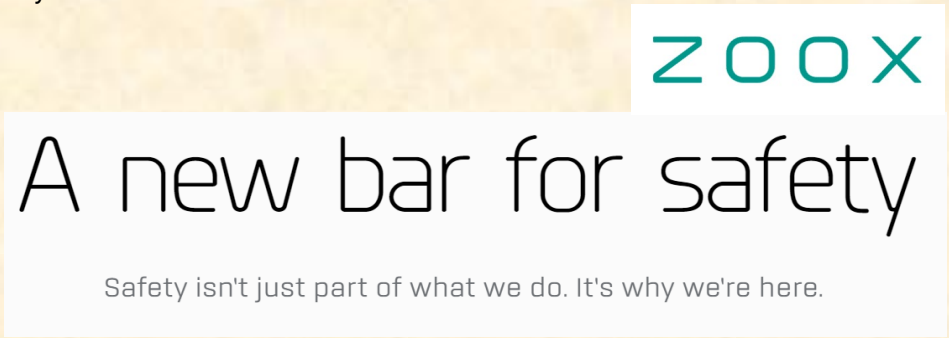
Carnegie Mellon University

https://waymo.com/safety/

https://getcruise.com/safety/



https://motional.com/safety-philosophy

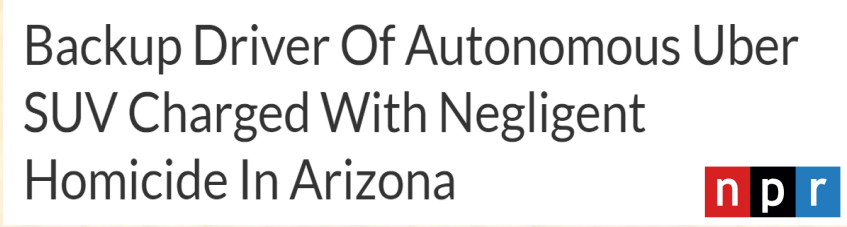


https://zoox.com/safety/

Automated vehicle Incidents

Carnegie Mellon University

- Uber ATG fatality, Tempe AZ/US: March 2018
  - Uber ATG closed: January 2021
- Local Motors shuttle driver injury
  - Company closed: Jan. 2022
- Pony.AI crash, CA/US: Oct. 2021
  - Uncrewed test permit revoked
- Easymile shuttle phantom braking injuries: (2019, 2020)
- Cruise & Waymo issues in San Francisco
  - Stalling in traffic, emergency responder issues; fire truck crash
- Cruise pedestrian dragging injury: Oct. 2023
  - Testing permits revoked; operational shutdown



2020 -- http://bit.ly/3Mwp1BG

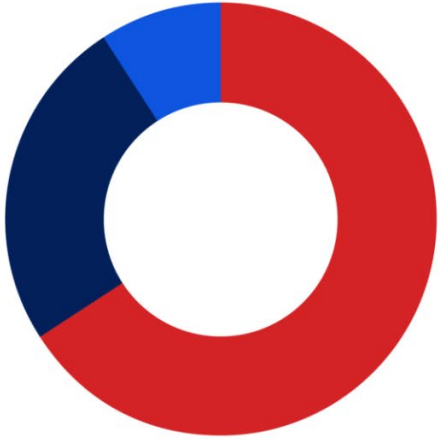
© 2024 Philip Koopman

Public Trust Is Eroding

Carnegie Mellon University

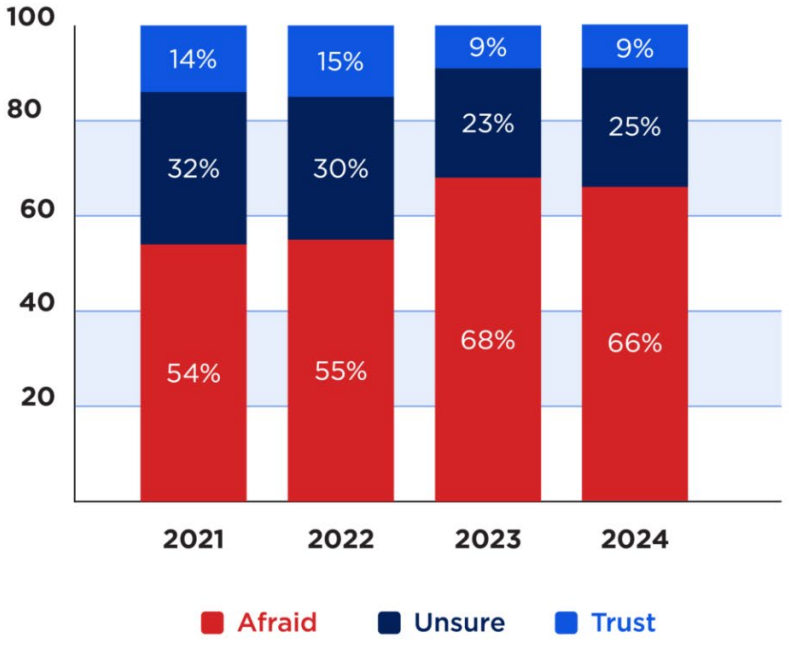
Driver Attitudes Toward Self-Driving Vehicles

2024 Survey Responses



Afraid	66%
Unsure	25%
Trust	9%

Driver Attitudes Over Time



[AAA: https://bit.ly/48YPgZe]

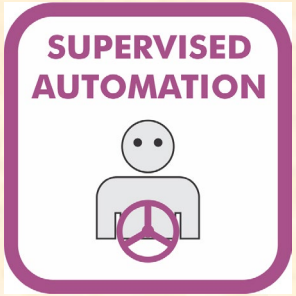
© 2024 Philip Koopman

Types of Vehicle Automation

Carnegie Mellon University

- Driver Assistance
  - The person drives; the car helps

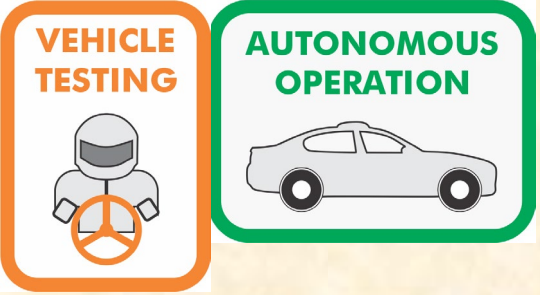
- Supervised Automation
  - The car mostly drives; the person helps
  - Lane Centering technology



- Autonomous
  - The car does all the driving

- Testing
  - Test driver compensates for automation defects

© 2024 Philip Koopman



Carnegie Mellon University

TRUE

You Can Ride in an Autonomous Vehicle Today

Robotaxi Deployments

Carnegie Mellon University

- Waymo:
  - Phoenix, San Francisco, Austin, Los Angeles
- Motional:
  - Las Vegas
- Cruise:
  - Paused (previously multiple cities)
- This will likely change over time
  - Other companies; other cities

[Waymo]

© 2024 Philip Koopman

Other Testing & Deployments

Carnegie Mellon University

- Other pilots/deployments/testing
  - Local parcel delivery
  - Low speed shuttles
  - Full size buses
  - Middle-mile trucks
- Driver-out operations over time
  - Varies by company, operational concept
- Chinese robotaxis [https://bit.ly/3TJ4Kw8]
  - Policy seems to be continuous remote safety supervision, for now

[Nuro]

© 2024 Philip Koopman

Remote Operators

Carnegie Mellon University

- Remote operator roles
  - Full remote driving
  - Remote safety operator
  - Remote intervention when requested
  - …
- Remote operator and safety
  - Infrequent remote interaction perhaps OK
    - Depends on the specifics
  - Can remote operator cause safety issues?
  - Can lack of remote operator request cause safety issues?
- Many open questions here…

This Cruise in San Francisco seemingly could not figure out how to pull aside on a narrow street to let a buss pass.

Cruise confirms robotaxis rely on human assistance every four to five miles

https://bit.ly/4apOeqc

© 2024 Philip Koopman

Autonomous Pilot Deployments Are Already On Public Roads; Testing Continues

Carnegie Mellon University

MYTH

Personally Owned Vehicles Can Drive Themselves Safely

Personal Vehicles Require Supervision

Carnegie Mellon University

- Personal vehicle driving automation:
  - Adaptive cruise control
  - Automated lane centering
- Driver plays a role in safety
  - *Limits to automation capabilities*

Culver City CA, 2018 [NTSB HAB-19/07]

- So-called “Level 2/2+” systems
  - Hands-on: Tesla, Audi, Kia, Mercedes Benz, Volvo, Nissan, Infiniti
  - Hands-free: GM, Ford, BMW [https://bit.ly/4ciSDx3]

- So-called “Level 3” systems
  - Mercedes Benz (but driver must still monitor traffic conditions)

© 2024 Philip Koopman

Things Can Go Very Wrong

Carnegie Mellon University

- Automation complacency:
  - Drivers over-trust automation
  - Attention wanders
  - Temptation to stop monitoring
- Bad things can happen very quickly
  - Delray Beach fatality, 2019
  - Engagement 9.9 secs before crash
  - No human steering for 7.7 seconds
- Driver Monitoring technology might help…
  … but is still a work in progress

Time to Collision = 1 second
Distance to Impact = 101 feet

Tractor-trailer combination vehicle still in motion and completely blocking all US 441 southbound lanes.

Delray Beach, FL, 2019
NTSB HAB-20/01

© 2024 Philip Koopman

IIHS: Only 1 of 14 Systems "Acceptable"

Carnegie Mellon University

	Overall rating	Driver monitoring	Attention reminders	Emergency procedures	Driver involvement: Lane change	Driver involvement: ACC resume	Driver involvement: Cooperative steering	Driver involvement: Safety features
Lexus Teammate with Advanced Drive 2022-24 Lexus LS	A	M	G	A	G	A	G	G
General Motors Super Cruise 2023-24 GMC Sierra	M	P	G	G	P	A	P	G
Nissan ProPILOT Assist with Navi-link 2023-24 Nissan Ariya	M	M	A	M	G	G	G	A

...

11 more

[IIHS 2024; Other 11 rated "Poor" https://www.iihs.org/ratings/partial-automation-safeguards]

© 2024 Philip Koopman

Automated Steering Requires Continuous Human Driver Attention – Not Really “Self-Driving”

Carnegie Mellon University

Misleading

People Are Inherently Terrible Drivers

…

It’s Complicated

The Myth of 94% Human Error

Carnegie Mellon University

- "94% of serious crashes are due to human error"
  - *– Consumer Technology Association*
    Testimony to US Congress, July 2023

[https://bit.ly/3TNMdi1]

- Humans failed to prevent ≠ human caused
  - What the NHTSA source study actually says:

"The critical reason was assigned to drivers in an estimated 2,046,000 crashes that comprise 94 percent of the NMVCCS crashes at the national level. [DOT HS 812 115]

However, in none of these cases was the assignment intended to blame the driver for causing the crash."

Benefits of Automation

SAFETY

The safety benefits of automated vehicles are paramount. Automated vehicles' potential to save lives and reduce injuries is rooted in one critical and tragic fact: 94 percent of serious crashes are due to human error. Automated vehicles have the potential to remove human error from the crash equation, which will help protect drivers and passengers, as well as bicyclists and pedestrians. When you consider more than 35,092 people died in motor vehicle-related crashes in the U.S. in 2015, you begin to grasp the lifesaving benefits of driver assistance technologies.

https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety

© 2024 Philip Koopman

Jan. 2022:

‘It Ain’t 94 Percent’: NTSB Chair Jennifer Homendy Discusses the Role of Human Error in Car Crashes

6:01 PM EST on January 31, 2022

https://bit.ly/493OUjX

Industry: Replace Terrible Human Drivers

Carnegie Mellon University

Humans are terrible drivers

42,795 Americans were killed in car crashes last year

You might be a good driver, but many of us aren't.
People cause millions of accidents every year in the US.
Cruise driverless cars are designed to save lives.
Our cars were involved in 92% fewer collisions as the primary contributor.* They also never drive distracted, drowsy or drunk.

Kyle Vogt
@kvogt

Follow

We ran this full-page ad in @nytimes and several local papers today.

Human drivers aren't good enough. America can do better, and it is time we fully embrace AVs.

Last edited 11:45 AM · Jul 13, 2023 · **956K** Views

July 2023

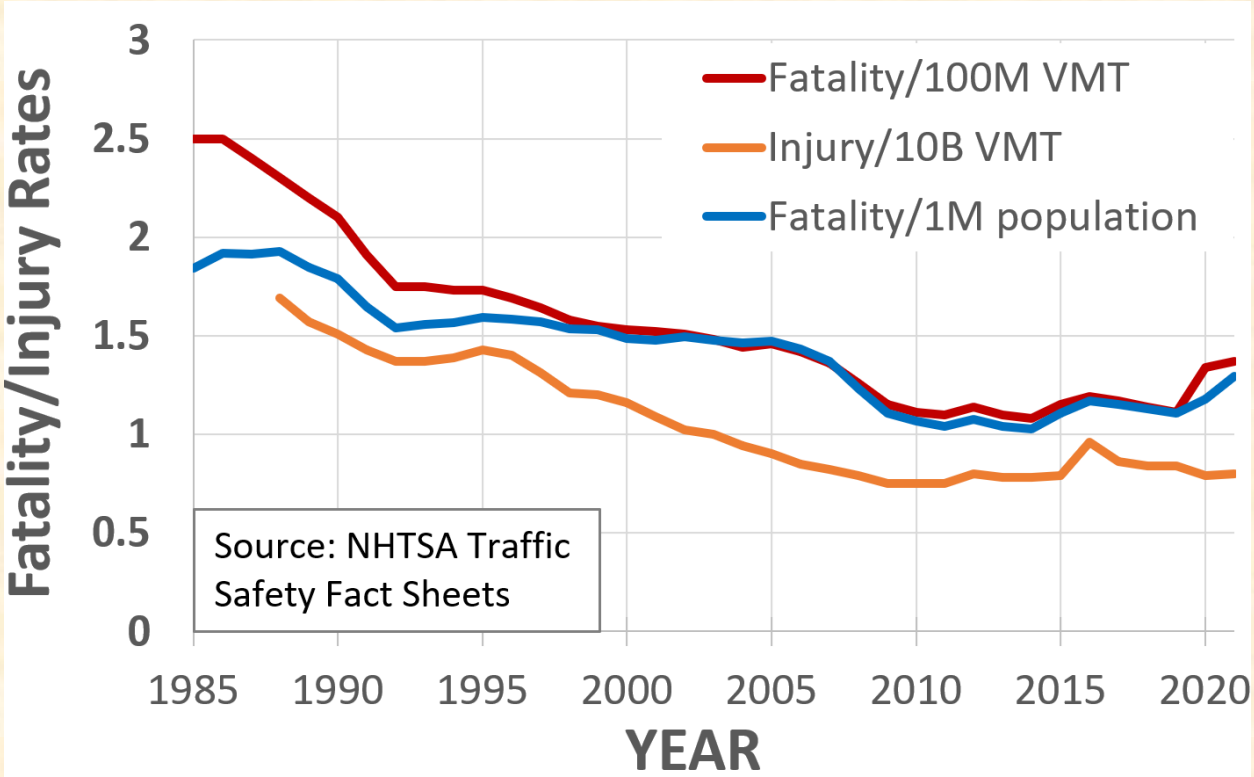
© 2024 Philip Koopman

Human Drivers Can Improve

Carnegie Mellon University

- Fatality/injury rate reduced:
  - Fatality/VMT: 60%
  - Injury/VMT 47%
  - Fatality/Person 67%

Multiple factors at work to improve safety



© 2024 Philip Koopman

Might We Do Better?

Carnegie Mellon University

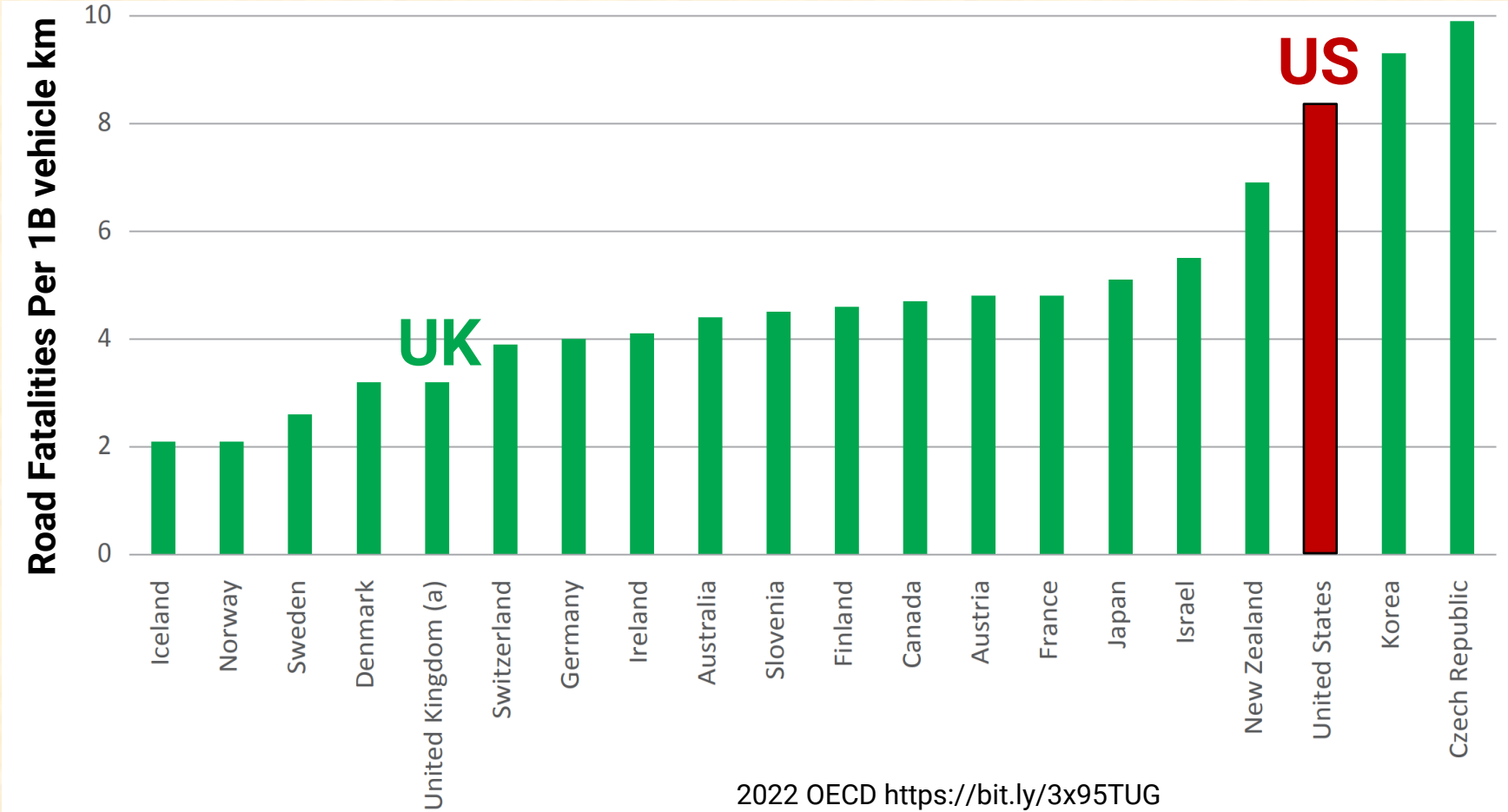
- Alcohol-related road fatalities:
  - US: 1985: 41% of fatalities
    2019: 28% of fatalities [NHTSA Traffic Facts]
  - UK: 1985: 18% of fatalities https://bit.ly/4cprcS2
    2019: 13% of fatalities https://bit.ly/3Tspve2

- US fatality rates: 1985 2.50 /100M VMT [NHTSA]
  2019 1.11 /100M VMT (1.37 in 2021)
- UK fatality rates: 1985 2.67 /100M VMT [dft.uk.gov]
  2019 0.51 /100M VMT (0.52 in 2021)

© 2024 Philip Koopman

Many Countries Do Better Than the US

Carnegie Mellon University



2022 OECD https://bit.ly/3x95TUG

Better Road Safety Does Not Require Using Computer Drivers

Carnegie Mellon University

TRUE

Computer-Controlled Active Safety Features Can Improve Safety

Active Safety Can Really Work!

Carnegie Mellon University

- Example Warning features:
  - Back-up camera & warning
  - Tire pressure monitoring
  - Rear cross-traffic alert

Automatic emergency braking

[IIHS: https://bit.ly/3PrXUZa]

⬇ 50%	Front-to-rear crashes
⬇ 56%	Front-to-rear crashes with injuries
⬇ 14%	Claim rates for damage to other vehicles
⬇ 24%	Claim rates for injuries to people in other vehicles
⬇ 41%	Large truck front-to-rear crashes

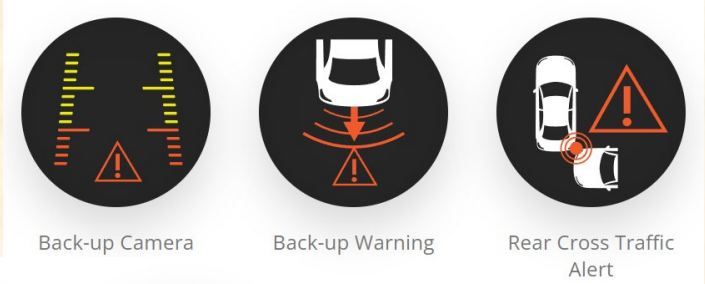
- Example Active Safety:
  - Electronic Stability Control (ESC)
  - Automatic/Advanced Emergency Braking (AEB)
  - Lane Keeping Assistance (LKA)
    - *Momentary* nudge at lane boundary
  - *Does NOT INCLUDE* sustained steering (Lane Centering)

© 2024 Philip Koopman

Example Car Safety Features

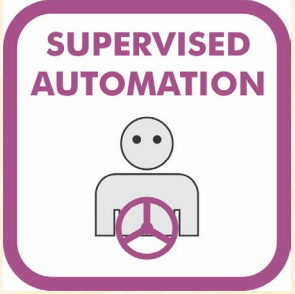
Carnegie Mellon University

- http://MyCarDoesWhat.org
  - List, icons & descriptions



© 2024 Philip Koopman

Computer-Controlled Features Can Improve Safety



Carnegie Mellon University

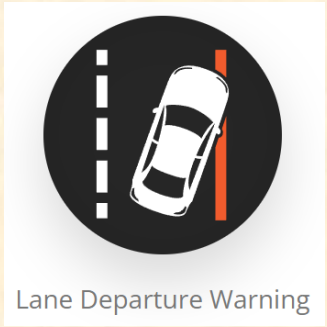
MYTH

Automated Steering Improves Driving Safety

Automated Steering Vs. Active Safety

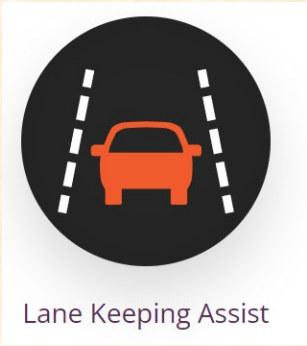
Carnegie Mellon University

- Active safety:
  - Lane Departure Warning (LDW)
  - Lane Keeping Assist (LKA)
    - Momentary nudge at lane boundary



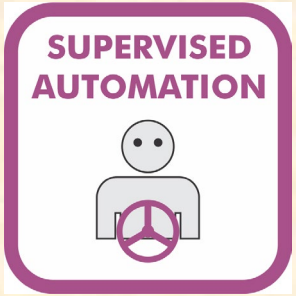
Lane Departure Warning

[MyCarDoesWhat.org]



Lane Keeping Assist

- Automated steering:
  - Lane Centering/Autosteer
    - Sustained steering control

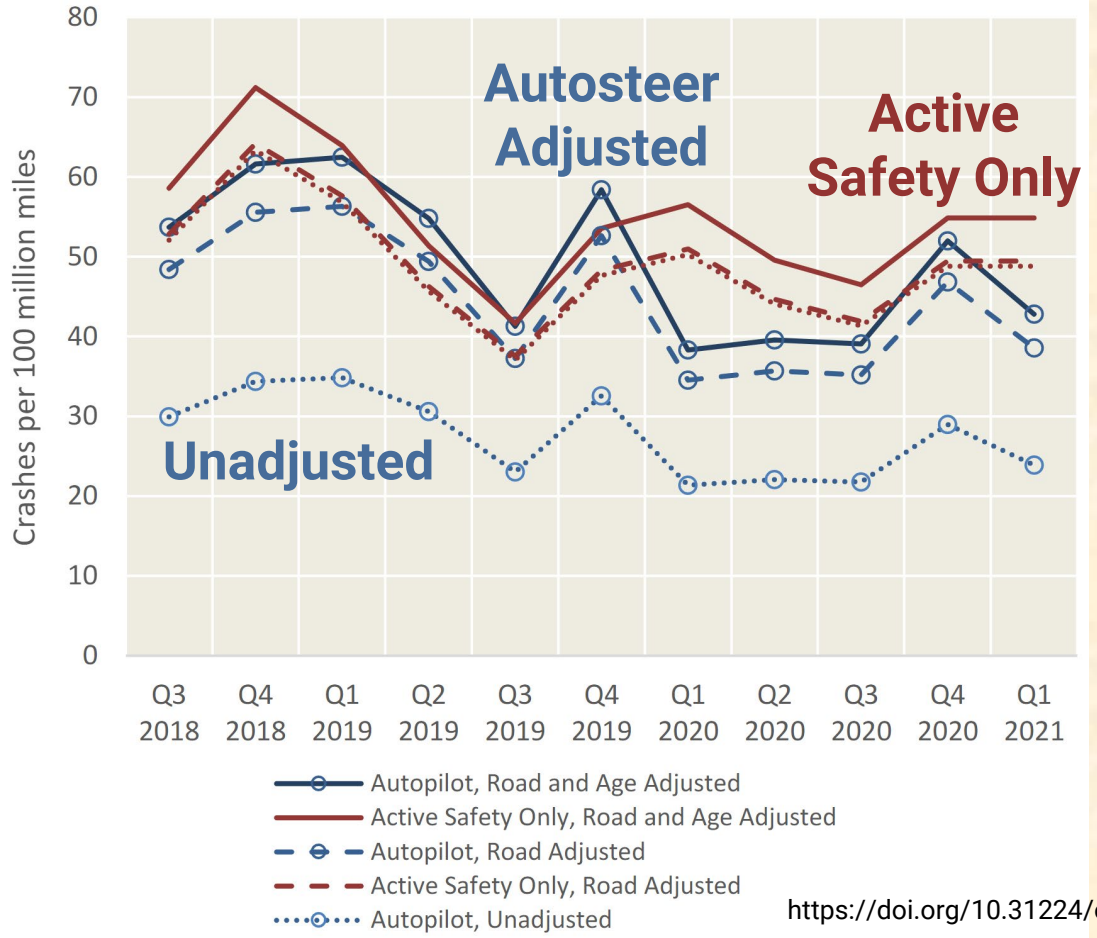


  - It's not really "assist" – it is actually steering the vehicle
    - Driver is no longer continuously controlling vehicle
    - For decades we've known this causes "driver drop-out" attention loss

© 2024 Philip Koopman

Active Safety Makes The Difference

Carnegie Mellon University



- Noah Goodall, 2021
  - Analyzed the data
- Claimed safety benefits diminish adjusted for:
  - Active safety feature benefits
  - Driver age
  - Freeway vs. other roads

https://doi.org/10.31224/osf.io/m8j6g

© 2024 Philip Koopman

Automated Steering Not A Safety Feature

Carnegie Mellon University

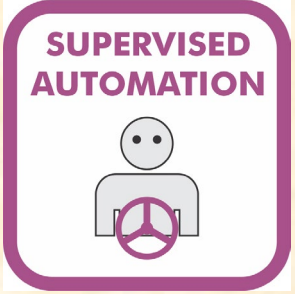
- 2024: Insurance Institute for Highway Safety (IIHS)

Safety features

There is little evidence that partial automation has any safety benefits, so it's essential that these systems can only be used when proven safety features are engaged. These include seat belts, AEB and lane departure prevention.

IIHS: March 2024
https://bit.ly/3Vsi35k

Automated Steering Is A Convenience Feature, Not A Safety Feature



Carnegie Mellon University

TRUE

People Are Terrible At Supervising Automation

Automation Bias & Complacency

Carnegie Mellon University

- Automation Bias
  - People tend to over-trust automated decision making
- Automation Complacency
  - Inattention to potential malfunctions
- Skill Degradation
  - Relying on automation degrades skills

See: https://en.wikipedia.org/wiki/Automation_bias

© 2024 Philip Koopman

NTSB Recommendations

Carnegie Mellon University

- NTSB H-17-41:
  - Incorporate system safeguards that limit the use of automated vehicle control systems to those conditions for which they were designed.
- NTSB H-17-42
  - Develop applications to more effectively sense the driver's level of engagement and alert the driver when engagement is lacking while automated vehicle control systems are in use.

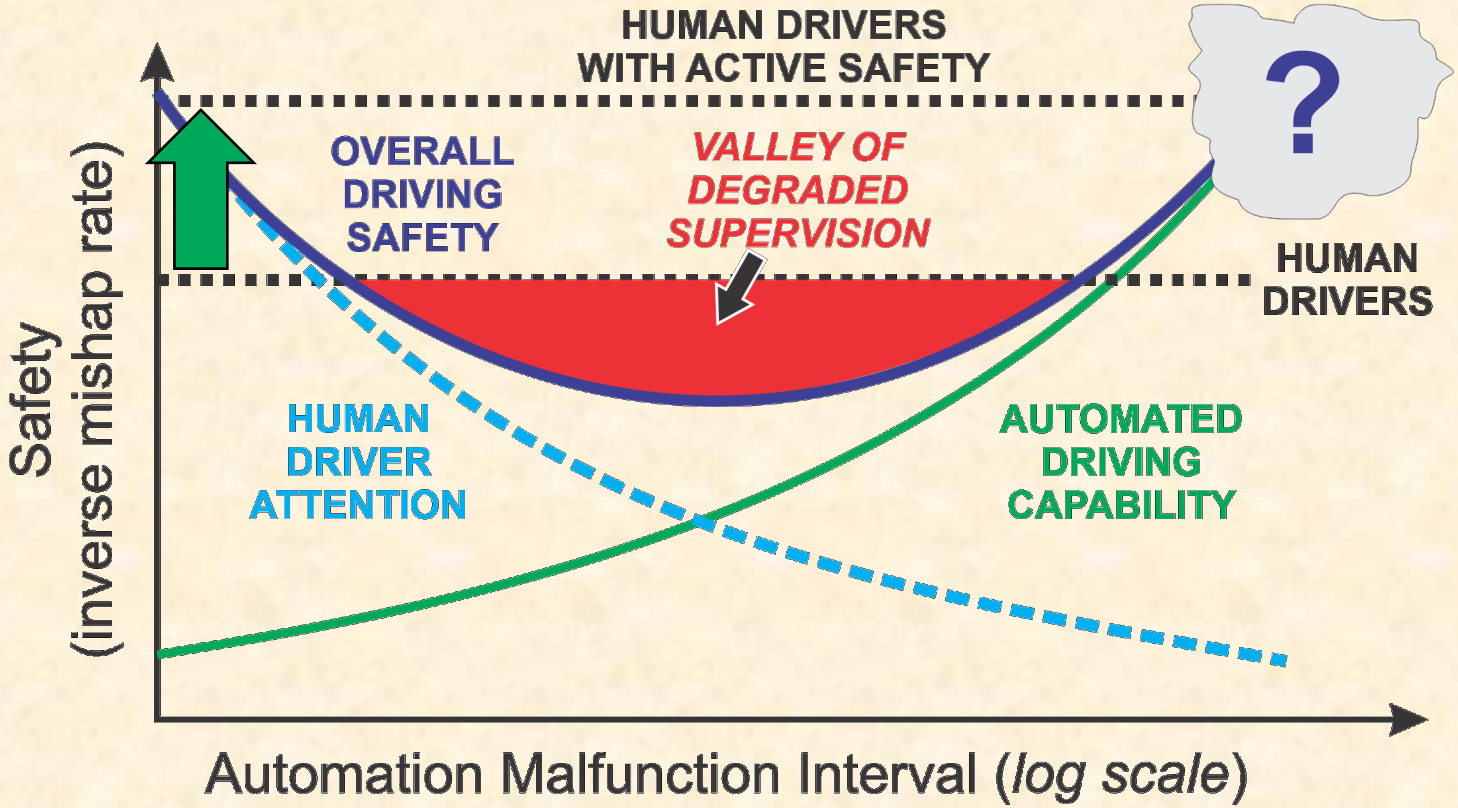
Figure 4: Florida Highway Patrol post-collision photograph depicting collision damage to Tesla. The yellow arrow indicates upper surface damage to the hood, while the red arrow indicates the pole impact damage. The rear window frame and trunk deck combination has been returned to the upright position.

Williston FL, May 2016
Fatality NTSB HAR-17/02

Also: H-17-37, H-17-38, H-17-39, H-17-40, H-17-43, H-20-2, H-20-3, H-20-4

© 2024 Philip Koopman

Risk of Degraded Safety



Shape of curves will vary by system & operational concept

© 2024 Philip Koopman

Driver Monitoring To The Rescue?

Carnegie Mellon University

- Driver Monitoring Technology
  - Steering wheel touch sensor
  - Face & gaze camera
  - Hand position sensing
  - …
- Some challenges:
  - Sensing challenges: darkness, sunglasses, gloves
  - Intentional misuse/abuse: covered camera, wheel weight
  - Determining mental state from a person's external features
  - What if monitoring shows drivers are unable to remain attentive?
    - The real challenge is driver attention management


[Euro NCAP]

© 2024 Philip Koopman

Driver Attention Management Is An Open Challenge

Carnegie Mellon University

MYTH

Ordinary Drivers Are Qualified To Test Driving Automation

Public Road Beta Testing

Carnegie Mellon University

- **Beta Testing: Operation in intended environment**
  - **Expectation that software can/will have defects**

Warning [Full Self-Driving (Beta) Tesla Owner Manual]

Model S may quickly and suddenly make unexpected maneuvers or mistakes that require immediate driver intervention.

Full Self-Driving (Beta)

Tesla 2023.44.30.7 Release Notes

Last updated 23-Mar-2024

You can enable Full Self-Driving (Beta) by tapping 'Controls' > 'Autopilot' > 'Full Self-Driving (Beta)' and following the instructions.

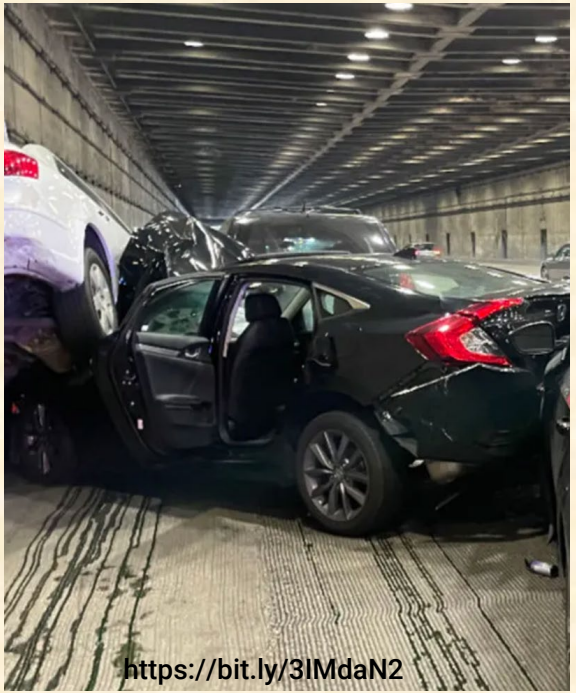
Full Self-Driving is in early limited access Beta and must be used with additional caution. It may do the wrong thing at the worst time, so you must always keep your hands on the wheel and pay extra attention to the road. Do not become complacent. When Full Self-Driving is enabled your vehicle will make lane changes off highway, select forks to follow your navigation route, navigate around other vehicles and objects, and make left and right turns. Use Full Self-Driving in limited Beta only if you will pay constant attention to the road, and be prepared to act immediately, especially around blind corners, crossing intersections, and in narrow driving situations.

https://bit.ly/3vzWllc

Road Testing Can Cause Real Harm

Carnegie Mellon University

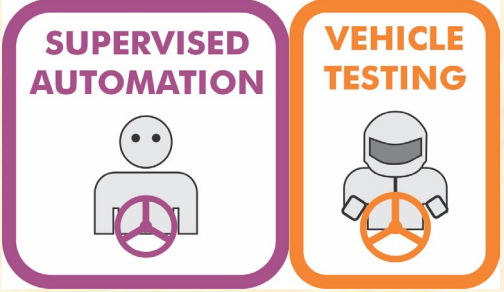
- Safety testing:
  - Does intended things correctly
  - Does not have unsafe surprises
  - *Testers face risk of dangerous misbehaviors*
- Accepted industry practices
  - Simulations & test track before road test
  - Testers must have special training
  - Testing per test plan; avoid known defects
- Ordinary retail customers should never perform the role of “tester”



SF Bay Bridge Beta multi-injury Testing Crash, Nov. 24, 2022

© 2024 Philip Koopman

Customers Cosplaying "Beta Tester" Expose Everyone To Undue Risk



Carnegie Mellon University

TRUE

Blaming Drivers Deflects Accountability Away From Companies

The Moral Crumple Zone

Carnegie Mellon University

- Moral Crumple Zone Strategy:
  - Human operator is a system component to bear the brunt of moral & legal responsibility

1. Design a known unsafe system
2. Deploy with a human operator
3. System fails due to safety defect
4. Blame the human operator
5. Scrutiny deflected from defect; safety defect is not corrected

Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction (pre-print)

Engaging Science, Technology, and Society (pre-print)

29 Pages
Posted: 3 Apr 2016
Last revised: 15 Mar 2019

Madeleine Clare Elish
Google Inc.; University of Oxford - Oxford Internet Institute

Date Written: March 1, 2019

https://bit.ly/3x8bxGK

Autonomous Vehicle Tester Story

Carnegie Mellon University

- March 2018 Uber ATG Fatality
  - Pedestrian killed during testing in Phoenix AZ
- Complicated situation
  - Pressure to test aggressively
  - Controversy over driver behavior
- Operator faced criminal trial
  - Plea deal to undesignated felony (probation)
- Uber ATG faced no charges
  - Embarked on a safety path

'I'm the Operator': The Aftermath of a Self-Driving Tragedy WIRED MAR 8, 2022

https://bit.ly/3VrqnlZ

https://bit.ly/43IcfXH

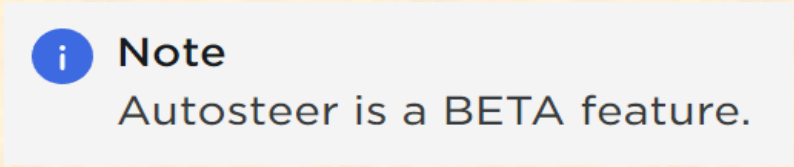
© 2024 Philip Koopman

Tesla Autopilot Double Fatality

Carnegie Mellon University

- Dec. 2019: Drove 74 mph through red light
  - Off-duty limousine driver using Autopilot
  - Ran red light after end of freeway
  - Killed two people in another vehicle
- Tesla faced no charges
  - Does not enforce highway-only
- Driver faced criminal trial
  - Plead no contest to vehicular manslaughter with gross negligence (probation)
- No apparent industry change

Tesla: https://bit.ly/3vndQVT



© 2024 Philip Koopman

Blaming Drivers Protects The Company, Not Necessarily Other Road Users

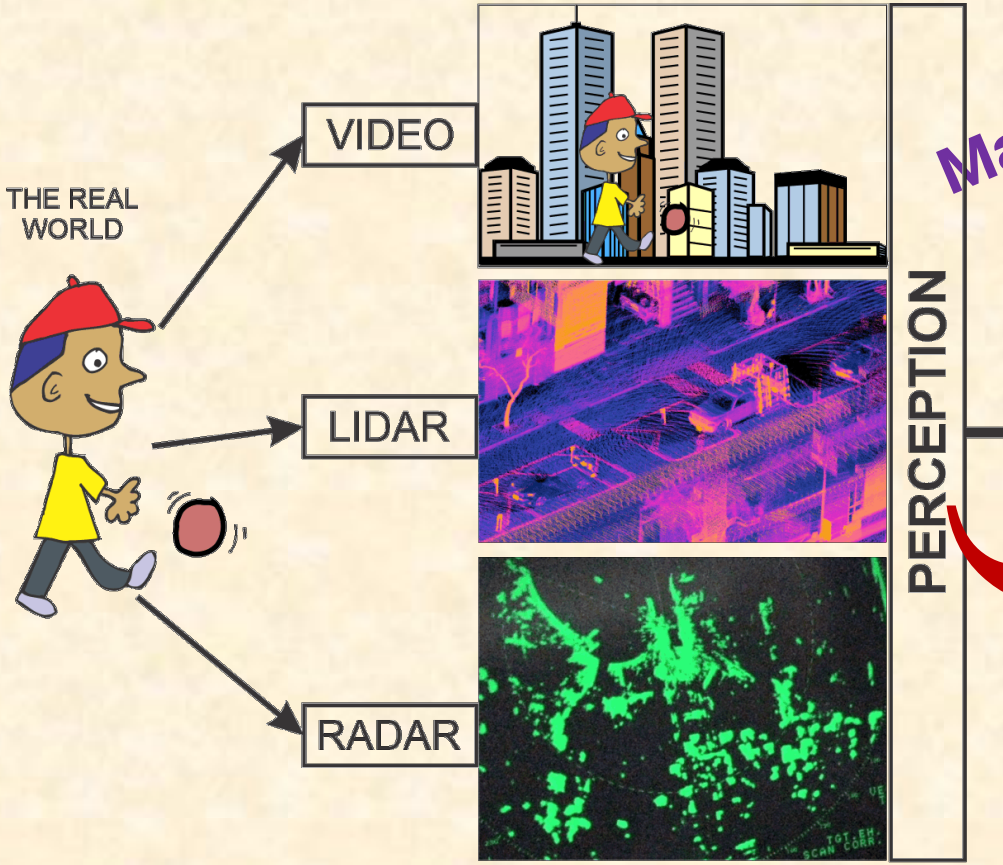
Carnegie Mellon University

MYTH

Lots Of Sensors Means No Avoidable Crashes

Perception Builds the World Model

Carnegie Mellon University



Machine Learning

COMPUTER'S WORLD MODEL:
"Child chasing ball into street 10 meters ahead"

→ **Path Planning & Motion Control**

Perception & prediction present a uniquely difficult assurance challenge

© 2024 Philip Koopman

Sensors Alone Do Not Ensure Safety

Carnegie Mellon University

"We're safe because we have LOTS of sensors!"

- Sensor fusion
  - What if sensors disagree?
- Perception/Prediction
  - What if system mis-classifies an object?
  - What if system mis-predicts object behavior?
- What if there is a planning/control fault?
  - March 2023: Robotaxi hits bus
    - Detected back half of articulated bus
    - Decided to consider only front half in planning
  - April 2023: recall for software defect

© 2024 Philip Koopman

Sensors Aren't Enough; Perception And Prediction Are Critical for Safety

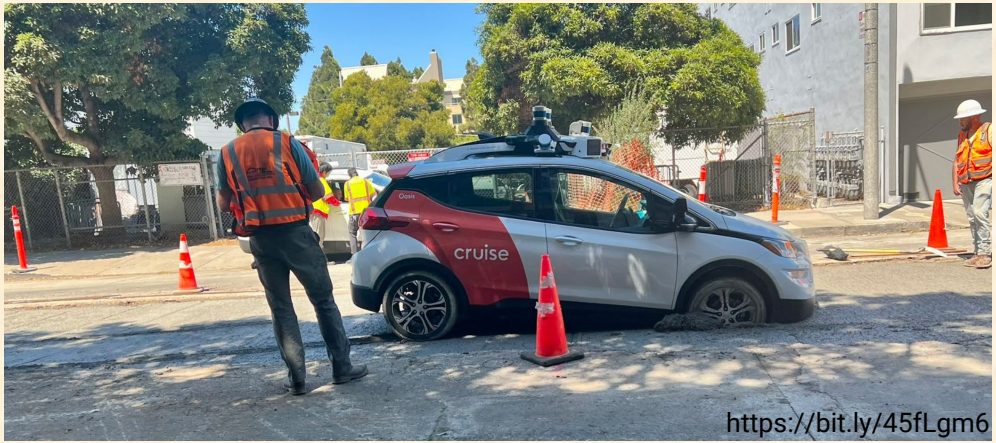
Carnegie Mellon University

Misleading

Computers
Won't Drive Drunk

Human Error → Robot Error

Carnegie Mellon University



https://bit.ly/45fLgm6

August 2023: Driving into Wet Concrete

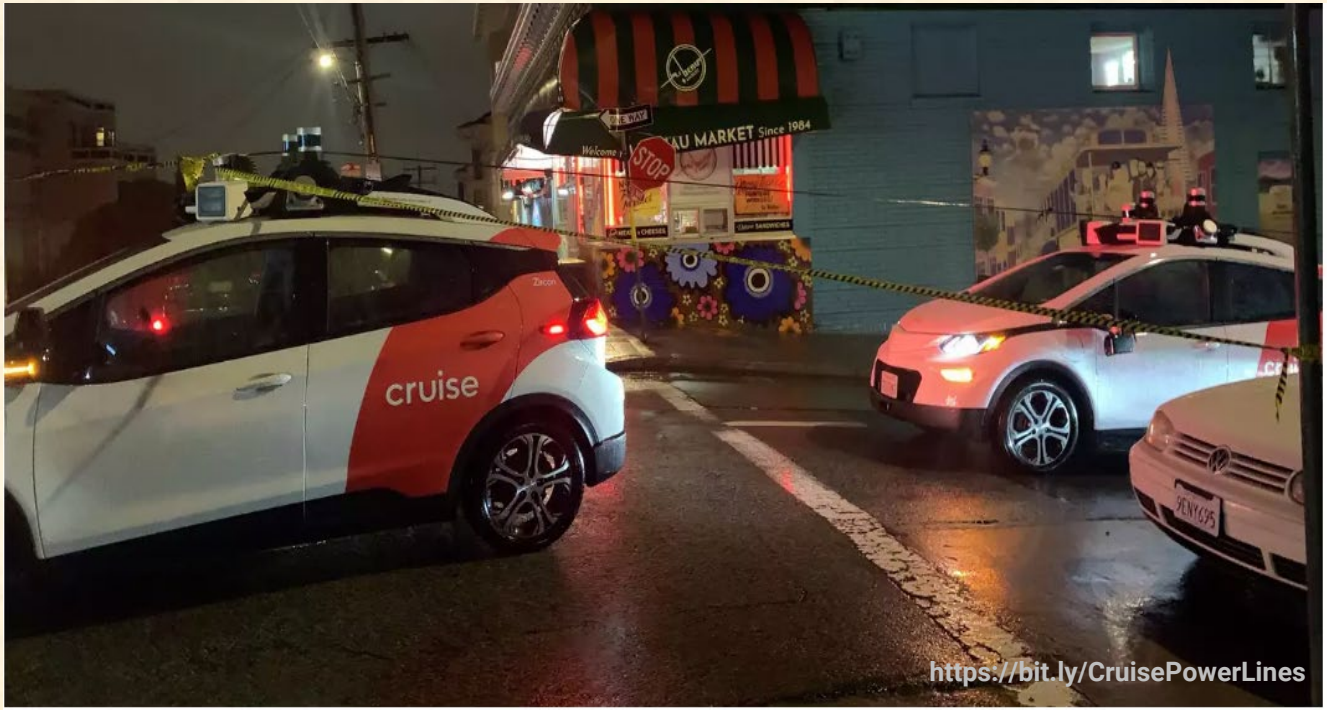
Aug. 2023:
Injury crash with fire truck.
CA DMV asked Cruise to cut active fleet size in half.

https://bit.ly/49POy27

© 2024 Philip Koopman

Handling Non-Crash Hazards

Carnegie Mellon University



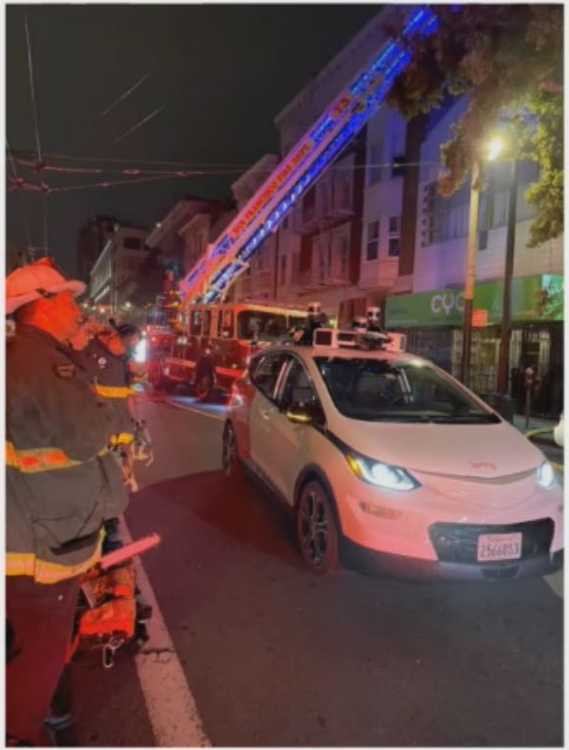
Two Cruise cars in San Francisco became wrapped in downed Muni wires and caution tape at Leavenworth Street and Clay Street on March 21, 2022.

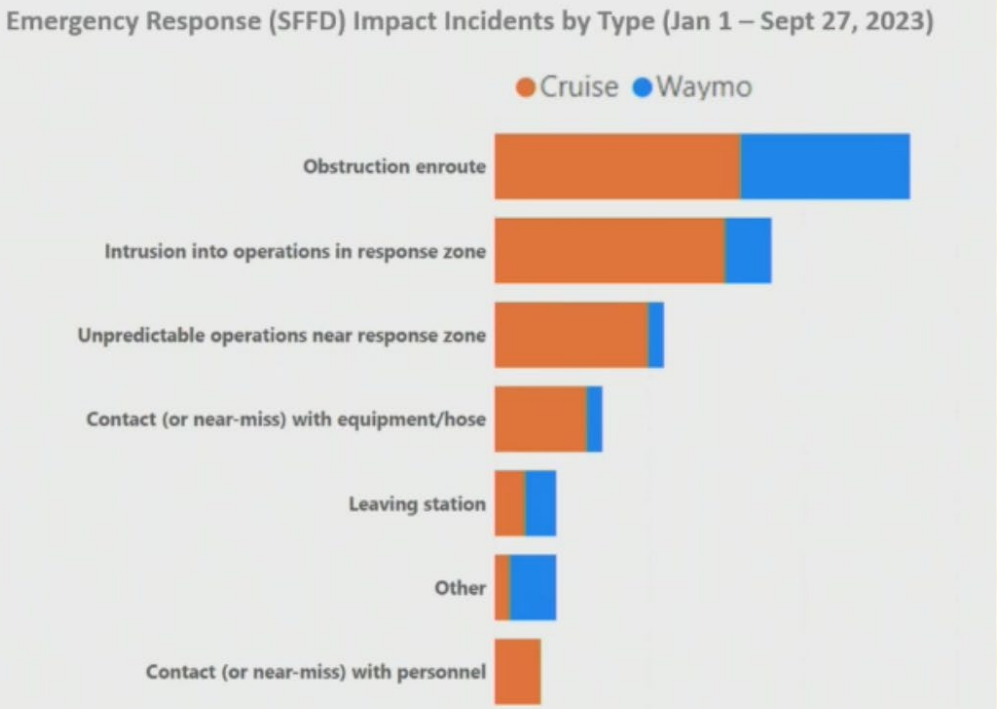
Courtesy of John-Phillip Bettencourt

City of San Francisco Concerns

Carnegie Mellon University

AV driving that interferes with emergency response





https://bit.ly/41cwJGI

Beyond Just Avoiding Crashes

Carnegie Mellon University

- Human drivers are imperfect
  - Drunk, DUI, tired
  - Aggressively violate road rules
- Robot drivers are imperfect
  - Software defects
  - Challenged by subtle context
  - Challenged by rare events
  - Errors in building model of the external world
  - Potential errors by remote human operators

August 2023
Nobody was hurt.
Does that make this safe?

© 2024 Philip Koopman

Robot Drivers Will Fail – Sometimes Differently Than Human Drivers

Carnegie Mellon University

TRUE

Safe Enough Requires More Than "Safer Than Human Driver"

What People Mean By "Safe"

Carnegie Mellon University

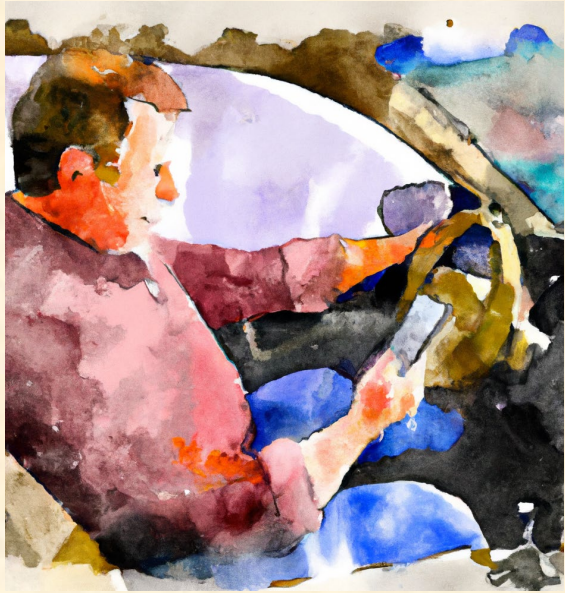
- Human drivers are bad, so computers will be safe
- "Safety is our #1 priority"
- Safe driving behavior / roadmanship
- Tested/simulated for millions of miles
- Risk is managed via insurance
- Conforms to safety standards
- Safety cases supported by evidence
- Positive Risk Balance (better than human)

[Dall-e]

© 2024 Philip Koopman

Positive Risk Balance

- Positive Risk Balance: safer than a human driver
- But which human driver?
  - 28% Alcohol/driving under influence fatalities
  - 26% speed-related, 9% distracted, 2% drowsy
  - 60 year old driver is ~3.5x better than 16 y.o.
- Where/Who?
  - 3.4x fatality per VMT variation by US state
  - Victim demographic (e.g., pedestrians)
- Which vehicle?
  - New cars have active safety – BUT average car age ~12 years



[Dall-e]

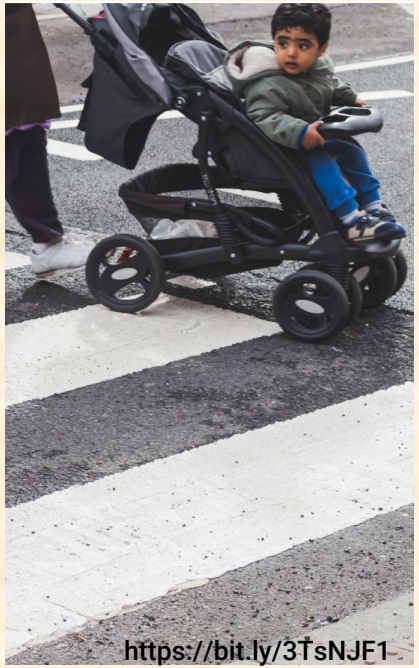
[DOT HS 813 060 & DOT HS 813 021] [AAA] [IIHS Fatality Fact Sheets State by State] [DOT HS 813 060]

© 2024 Philip Koopman

Other Safety Considerations

Carnegie Mellon University

- Avoid risk transfer to vulnerable populations
  - What if vulnerable road user risk increases?
- Avoid negligent driving behavior
  - What if breaking traffic rules leads to crashes?
- Fine-grain regulatory control of risks
  - Recalls due to specific risk, not net risk
- Ethical & equity concerns
  - What if some demographics are at increased risk?
- Potential for crash-by-crash comparison
  - What if “a human driver would never have made that mistake”?


https://bit.ly/3TsNJF1

© 2024 Philip Koopman

Need More Than Improved Statistical Average Safety

Carnegie Mellon University

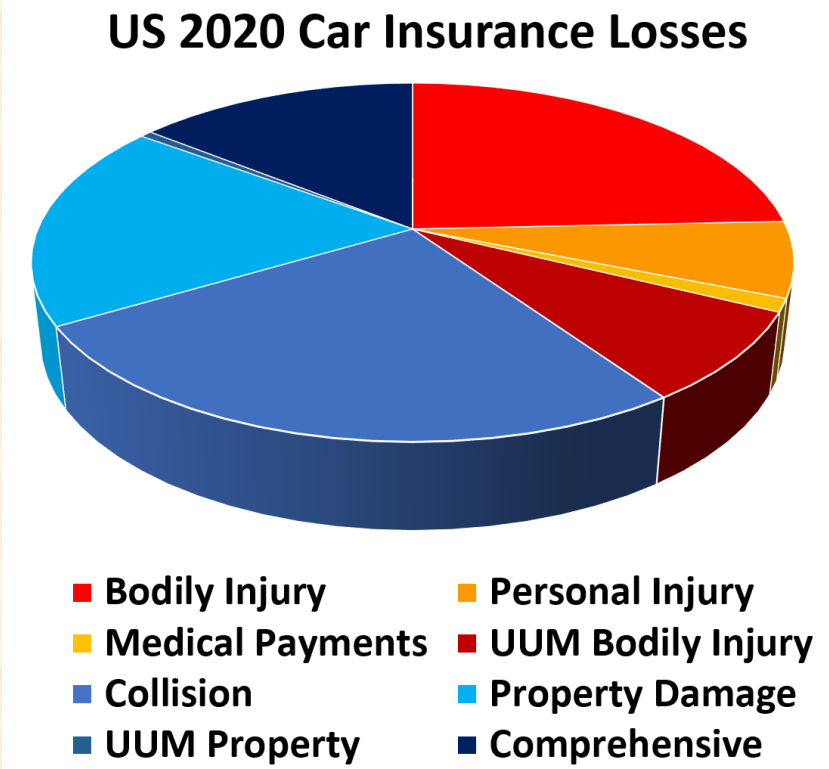
MYTH

Insurance Cost Pressure Will Ensure Acceptable Automated Vehicle Safety

Insurance Leverage for Safety

Carnegie Mellon University

- 2020 US Insurance Losses
  - Total $135B
  - 40% injury/medical losses
- 2020 Statistics
  - 2.9 Trillion vehicle miles
  - 267,585,097 Vehicles
  - 6,773,562 Collision Claims
  - 810,000 Vehicle Thefts
  - 38,824 Fatalities
    - Not all fatalities pay out big claims



[Data Source NAIC https://bit.ly/3TrWHm1]

© 2024 Philip Koopman

Affordable Insurance vs. Safety

Carnegie Mellon University

- "We are safe because we bought insurance"
  - Small numbers of vehicles limits exposure
  - Insurance company maximum payout: policy limit
- Affordable risk might exceed everyday safety
  - E.g., Life insurance for combat military personnel
- Insurance is about pricing risk, not ensuring safety
  - Customers pay for increased risk via premiums
  - Risk uncertainty perhaps more important to insurers

Affordable Insurance ≠ Acceptable Safety

https://bit.ly/46umY8J

© 2024 Philip Koopman

Net Risk Alone Is Not Safety

Carnegie Mellon University

- Redistribution of harm
  - What if more pedestrians, cyclists die?
  - What if more mishaps happen in historically disadvantaged areas?
- Negative risk externalities
  - Blocking fire trucks, ambulances
- What if known significant risks unmitigated?
  - Even if total fatalities decrease, is that OK?
- Fatalities due to breaking traffic rules
  - Humans break rules too…
    but they are *held accountable via negligence*

© 2024 Philip Koopman

Insurance Pressure Alone Will Not Ensure Acceptable Safety

Carnegie Mellon University

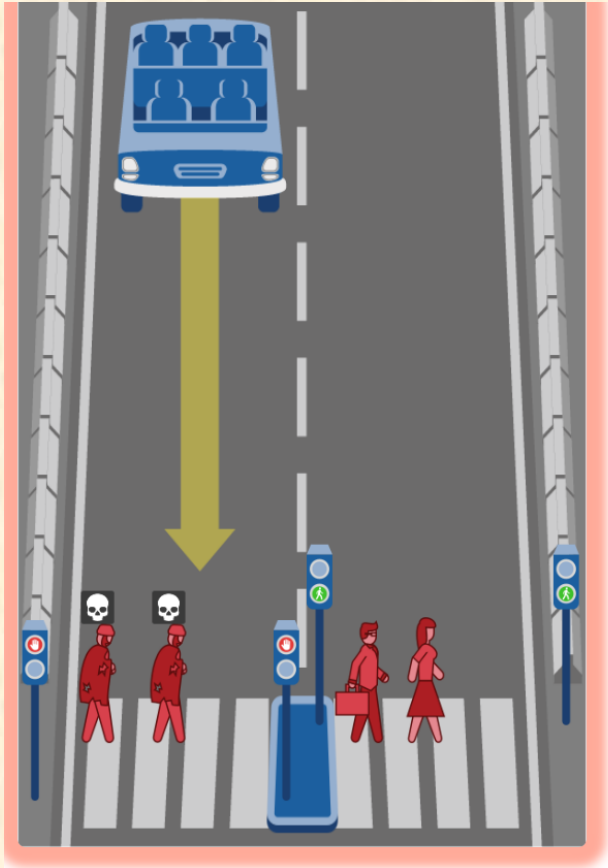
MYTH

Autonomous Vehicle Ethics Is All About The Trolley Problem

The Infamous Trolley Problem

Carnegie Mellon University

- Given a no-win situation, should the vehicle:
  - Kill 1 person to save 5?
  - Kill socially devalued people
    - Safety only for suit-wearers?
- This is a false dilemma!
  - How often will this happen?
  - Why was the car not equipped with redundant brakes?
  - Why did the car not roll itself over using a side barrier?



In this case, the self-driving car with sudden brake failure will continue ahead and drive through a pedestrian crossing ahead. This will result in ...
Dead:

- 2 homeless people

Note that the affected pedestrians are flouting the law by crossing on the red signal.

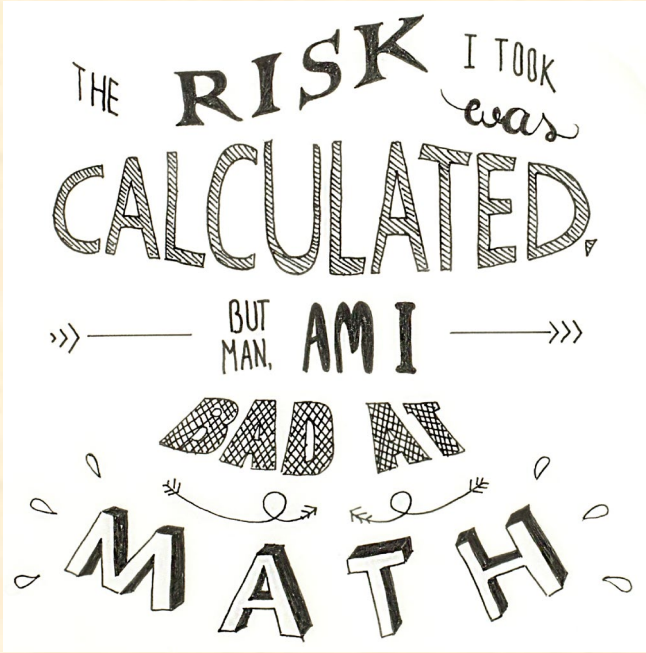
https://www.moralmachine.net/

© 2024 Philip Koopman

Ethics: Deployment Governance

Carnegie Mellon University

- #1 ethical issue is deployment governance
  - Who decides when to deploy based on what?
- Aggressive for-profit deployments
  - Existential financial & time pressure
  - Missing independent technical oversight
- Ethical deployment should address:
  - Publicly disclosed safety prediction
  - Inclusion of stakeholder concerns
  - Transparency of data & processes
  - Accountability for any losses

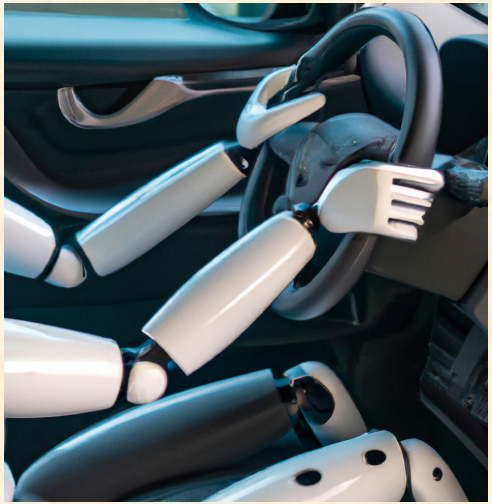


https://bit.ly/3rJeaJ4

© 2024 Philip Koopman

Equity Concerns

- Ride Hail made promises … with disappointing results
  - Why will for-profit robots turn out differently?
- Labor concerns:
  - Displaced ride-hail/taxi drivers
  - Displaced truck drivers
- Transportation access concerns:
  - Service for disabled in absence of regulations?
  - Cheap taxis undermining *safer* public transit
- Risk distribution concerns:
  - Testing risk might be imposed upon vulnerable people
  - Municipal preemption / no local control of issues



[Dall-e]

© 2024 Philip Koopman

Ethics/Equity Question: Who Decides What / When / Where To Deploy

Carnegie Mellon University

MYTH

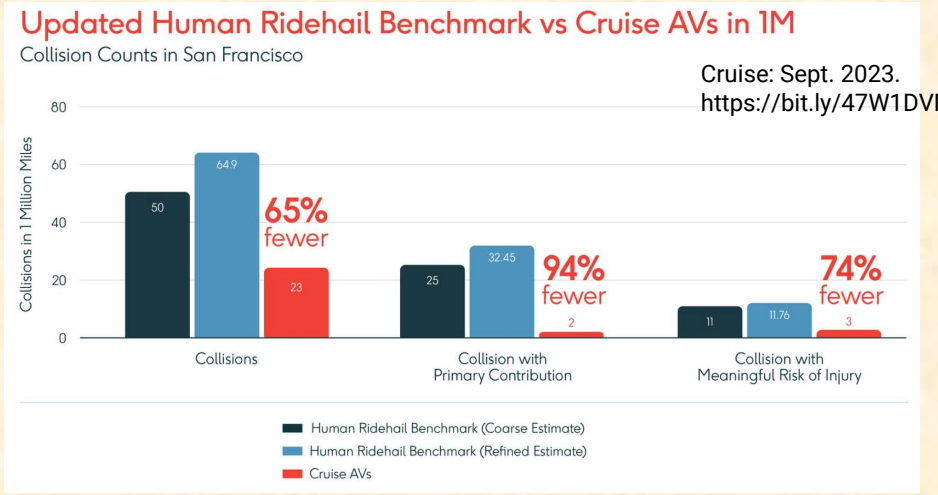
10 Million Good Miles Has Proven Autonomous Vehicles Are Safe

2023: Results From 1M+ Miles

Carnegie Mellon University

Waymo: Feb. 2023.
https://bit.ly/3N5F6xF

Emphasis on "at fault" crashes



Cruise: Sept. 2023.
https://bit.ly/47W1DVR

Sept. 2023 https://bit.ly/43KNmKZ
Waymo + Swiss Re Report
Based on 3.8 million miles

Waymo as of March 2024:
https://waymo.com/safety/

Waymo passenger injury August 2, 2023 -- the day after Swiss Re study decided to end data studied: https://bit.ly/47Z9pyb

© 2024 Philip Koopman

How Many Road Testing Miles?

Carnegie Mellon University

- Human driver miles per fatal crash: [NHTSA]
  - US: 1999: 98M VMT / 2021: 79M VMT
  - Includes drunk, impaired, speeding, …
- Statistically good as average human driver
  - 95% confidence
  - Need 237M – 294M VMT with no fatality
    - But at this point you likely have fatal crash(es)…
  - Rule of thumb: need 10x miles per crash
- Waymo 7.1M mile report: [Dec. 2023 at page 15; https://bit.ly/4cDuZvs]
  - "no statement…can be made" regarding serious injury/fatalities

Including test driver miles. Waymo as of March 2024: https://waymo.com/safety/

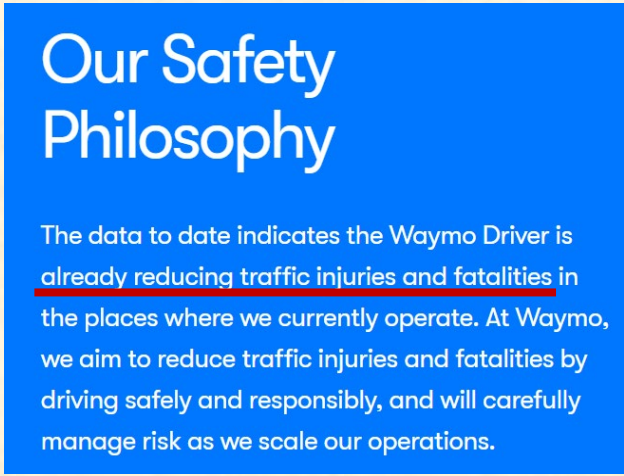
https://reliabilityanalyticstoolkit.appspot.com/mtbf_test_calculator

© 2024 Philip Koopman

Are Robotaxis Safer?

Carnegie Mellon University

- Robotaxi companies predict acceptable safety
  - Based on non-severe crash rates
  - With sometimes controversial limitations
  - Fatality & serious injury rates are predicted
- 300+ Million miles needed to confirm
  - Perhaps 5-10 million driverless miles now
  - With continually evolving software
  - Reduced fatality rates are still aspirational
- Declaring safety "victory" at this point is like claiming a medal … after the first mile in a marathon



[https://waymo.com/safety/]

© 2024 Philip Koopman

Companies Predict – But Cannot Yet Prove – Severe Injury/Fatality Safety

Carnegie Mellon University

MYTH

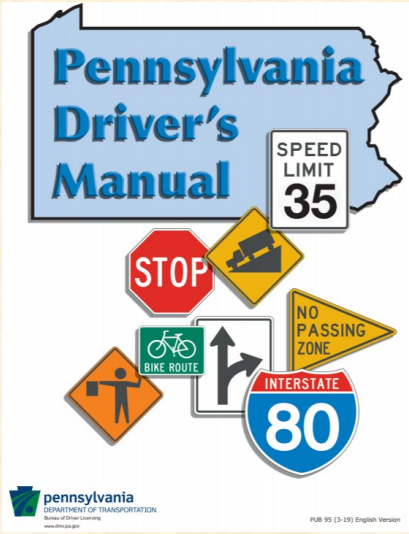
Road Testing Makes Autonomous Vehicles Safe

How About A Robot Driver Test?

Carnegie Mellon University

- Written test for Automated Driving System (ADS)
  - Does ADS know traffic laws & behaviors?
- Road test
  - Can ADS obey traffic laws?
  - Can ADS negotiate effectively with human drivers?
  - Can ADS resolve potentially ambiguous situations?
- Being a 16 year old human
  - How do we measure ADS judgment maturity?
  - Autonomous systems struggle with novelty, unknowns

➔ Need safety engineering, not just a driver test



© 2024 Philip Koopman

Brute Force Road Testing

Carnegie Mellon University

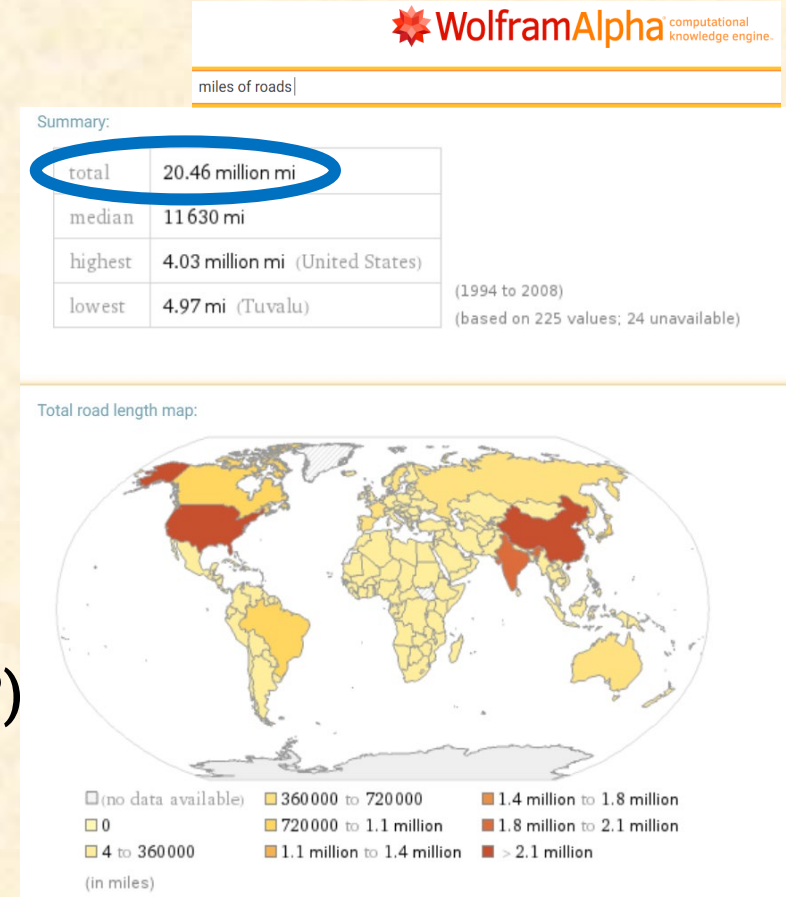
- If 100M miles/fatality…
  - Test 3x–10x longer than mishap rate
    ➔ Need 1 Billion miles of testing

- That's ~50 times on every road in the world
  - With fewer than 10 fatalities

…

  - Start over for each software update(?)

➔ Brute force testing impracticable



© 2024 Philip Koopman

The Challenge Is Covering Everything

Carnegie Mellon University

- Have you covered the possible unknowns?

https://goo.gl/J3SSyu

© 2024 Philip Koopman

Safety Requires an Accurate World Model

Carnegie Mellon University

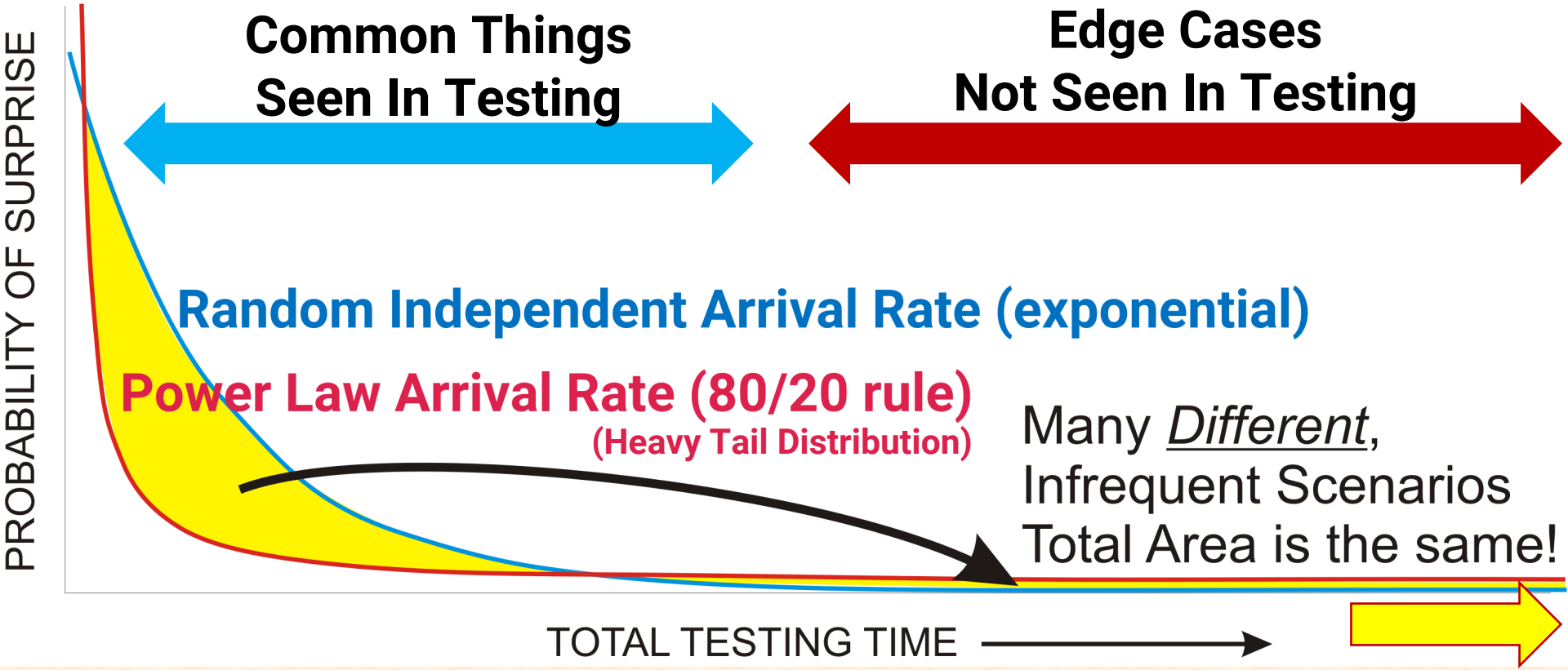
- Good prediction based on the world model
  - Classification accuracy affects prediction
  - Multiple possibilities for any object in any situation
- Safety limited by heavy tail scenarios (rare + important)
  - Probabilities of what happens next are context dependent
- Rare cases/unusual context can dominate safety

?

© 2024 Philip Koopman

Heavy Tail Distribution Of Surprises

Carnegie Mellon University



© 2024 Philip Koopman

Heavy Tail Edge Cases Explained

Carnegie Mellon University

- Where will you be after 1 Billion miles of testing?
- Assume 1 Million miles between unsafe "surprises"
  - Example #1:
    100 "surprises" @ 100M miles / surprise
    - All surprises seen about 10 times during testing
    - With luck, all bugs are fixed
  - Example #2: Heavy Tail
    100,000 "surprises" @ 100B miles / surprise
    - Only 1% of surprises seen during 1B mile testing
    - Bug fixes give no real improvement (1.01M miles / surprise)

https://goo.gl/3dzguf

Safety Engineering In A Nutshell

Carnegie Mellon University

- Safety Engineering Process
  - Identify hazards
  - Determine risk from hazards
  - Mitigate risk from hazards
  - Repeat until acceptable remaining risk


[Dall-e]

- Open challenges
  - How heavy tail is the distribution of event types?
  - Applying safety engineering to machine learning
  - How much/what type of remaining risk is acceptable?

Safety Depends On Engineering To Mitigate Rare, High-Consequence Events

Heavy-Tail Distribution Of Surprises Is A Challenge To Scalable Deployment

Carnegie Mellon University

MYTH

Safety Standards Don’t Exist and/or Would Stifle Innovation

Standards Set Expectation of Safety

Carnegie Mellon University

SYSTEM SAFETY	ANSI/UL 4600		**Safety Beyond Dynamic Driving**
DYNAMIC DRIVING FUNCTION	ISO 21448	SaFAD/ISO TR 4804	**Environment & Edge Cases**
FUNCTIONAL SAFETY	ISO 26262		**Equipment Faults**
CYBER-SECURITY	SAE J3061	SAE 21434	**Computer Security**
VEHICLE SAFETY	FMVSS	NCAP	**Basic Vehicle Functions**

REQUIRED

HIGHLY AUTOMATED VEHICLE SAFETY CASE

ANSI/UL 4600

ROAD TESTING SAFETY

SAE J3018

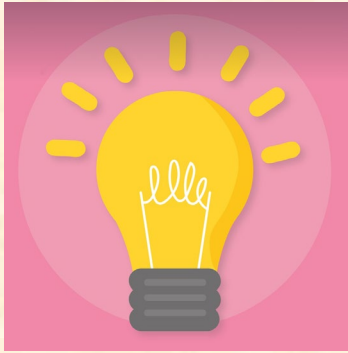
© 2024 Philip Koopman

Safety Standards & Innovation

Carnegie Mellon University

AV Industry: standards/regulation "Stifle Innovation"

- Do safety standards mandate particular technology?
  - NO – they require engineering rigor to show safety
- Do safety standards limit ability to test prototypes?
  - NO – primarily apply to public road deployment
- How do safety standards limit ability to road test?
  - Use of trained safety drivers and test plans
  - Big Red Button to disable computer control must actually work
- The burden for testing innovative approaches is minimal
  - Removing the safety driver is deployment, not safety testing



© 2024 Philip Koopman

Case Study: Loss of Titan Submersible

Carnegie Mellon University

OceanGate was also concerned that the classing process could slow down development and act as a drag on innovation. "Bringing an outside entity up to speed on every innovation before it is put into real-world testing is anathema to rapid innovation," it said.

In an interview with the Smithsonian magazine in 2019, Rush complained that the commercial sub industry had not "innovated or grown - because they have all these regulations".

The Guardian https://bit.ly/3PuM291

- Catastrophic 2023 implosion
  - Unorthodox construction techniques
  - Did not submit to external safety review
  - Developer attitude:
    - Real world testing is what matters
    - Regulation kills innovation

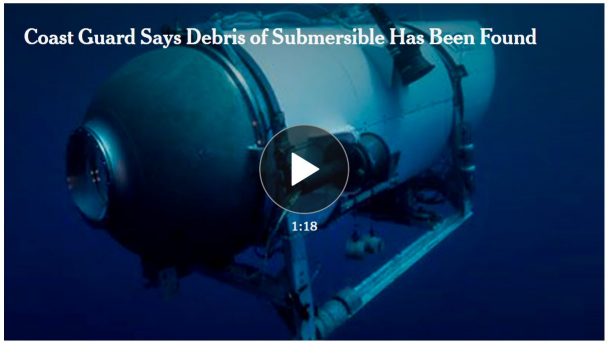
Missing Titanic Submersible

'Catastrophic Implosion' Likely Killed 5 Aboard Submersible

Pieces of the missing Titan vessel were found on the ocean floor, about 1,600 feet from the bow of the Titanic, the Coast Guard said. OceanGate Expeditions, the vessel's operator, said, "Our hearts are with these five souls."

Published June 22, 2023 Updated June 26, 2023

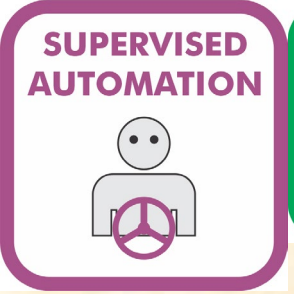
Share full article



The U.S. Coast Guard said parts of the Titan submersible found on the ocean floor indicate a "catastrophic implosion" of the vessel. OceanGate Expeditions, via Associated Press

© 2024 Philip Koopman

Safety Standards Deter UNSAFE Innovation



?????

Carnegie Mellon University

Government Regulation Will Ensure Safe Vehicle Automation

Robotaxi Regulatory System In Action

Carnegie Mellon University

Los Angeles Times

General Motors recalls all Cruise robotaxis after one dragged a pedestrian

General Motors is updating the software of its Cruise robotaxi vehicles after one struck and dragged a pedestrian in San Francisco last month, according to documents posted by safety regulators Wednesday. (Paul Sancya / Associated Press)

BY TOM KRISHER | ASSOCIATED PRESS NOV. 8, 2023 8:32 AM PT

- October 2, 2023 crash
  - Human-driven vehicle hits pedestrian
  - Cruise runs over person
  - Cruise robotaxi drags person *after* initial stop
  - Regulator interactions
  - Oct. 24, CA DMV suspends Cruise permits
  - Nov. 7, NHTSA Recall for post-collision response

© 2024 Philip Koopman

US Regulatory Posture

Carnegie Mellon University

- Federal / equipment safety: reactive (recalls)
  - NHTSA 2020 proposal to use industry standards stalled
  - Started collecting “SGO” crash data in 2021
- State / driver safety: administrative only
  - Texas, Arizona, etc. “open for business”
  - California: permits, licensing, reporting
    - But – impossible to ticket a robotaxi
- Municipal / adapt to locality: frustration
  - State preemption of localities
  - Pushback starting after San Francisco experiences

© 2024 Philip Koopman

Regulators Struggle with Novel Technology

Carnegie Mellon University

- Regulatory recalls
  - "Undue Risk" in the small – specific issues
  - Informed by test-centric standards
- Recalls historically specific, not net risk
  - Rolling through stop signs
  - Phantom braking
  - Malfunctioning display console
- Regulators struggling to predict safety outcomes in advance
  - Software safety & net risk are historically beyond regulatory scope

Part 573 Safety Recall Report

© 2024 Philip Koopman

Trend: System Safety Recalls

Carnegie Mellon University

- Feb 2022:

Tesla recall: 'Full Self-Driving' software runs stop signs

FILE - A 2021 Model 3 sedan sits in a near-empty lot at a Tesla dealership in Littleton, Colo. June 27, 2021. Tesla is recalling nearly 54,000 vehicles because their "Full Self-Driving" software lets them roll through stop signs without coming to a complete halt. Documents posted Tuesday, Feb. 1, 2022, by U.S. safety regulators say that Tesla will disable the feature with an over-the-internet software update. (AP Photo/David Zalubowski, File)

https://bit.ly/43xeX27

- Feb 2024:

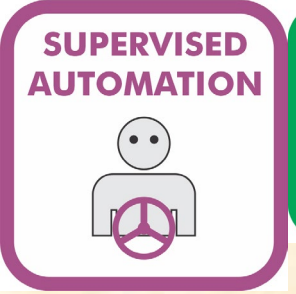
CR's Extensive Testing Shows That Tesla's Autopilot Recall Fix Does Not Address Safety Problems

The changes to warning messages and controls don't go far enough to prevent misuse and distraction, CR's car safety experts say

CR's Tesla Model Y
Photo: John Powers/Consumer Reports

https://bit.ly/3voV9B8

Federal Recall-Based Strategy Struggling To Deal With System-Level Safety



Carnegie Mellon University

MYTH

Product Liability Will Ensure Safe Vehicle Automation

Product Liability Is Not Enough

Carnegie Mellon University

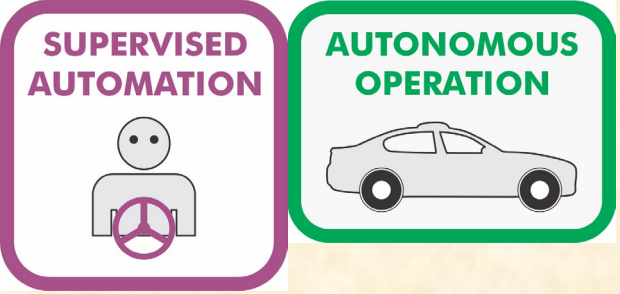
- Manufacturers are pushing for only product liability
  - Manufacturing defect, design defect, etc.
  - Must prove product presents undue risk
- Difficult and expensive to prove
  - Source code analysis expensive + painful
  - Class action requires commonality
    - With weekly neural network updates?
  - Poor machine learning explainability?
- Does this make sense if the car ran a red light and crashed?

Mercedes To Accept Liability When Autonomous Drive Pilot Is Engaged

Drive Pilot is a Level 3 system, and Mercedes will be the first automaker to accept legal responsibility when such a system is active.

© 2024 Philip Koopman

Product Liability Is The Wrong Tool For Most Automated Vehicle Crashes



Carnegie Mellon University

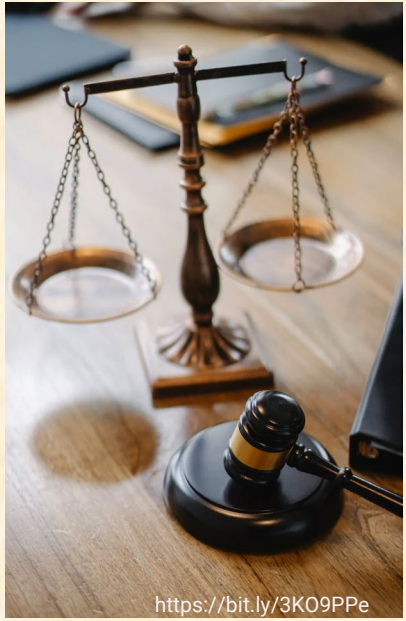
MYTH

Current Tort Liability Rules Will Ensure Safe Vehicle Automation

Tort Law for Non-Specialists

Carnegie Mellon University

- Civil Tort Law
  - Compensate a claimant who has suffered loss …
    proximately caused by …
    the negligence of another party.



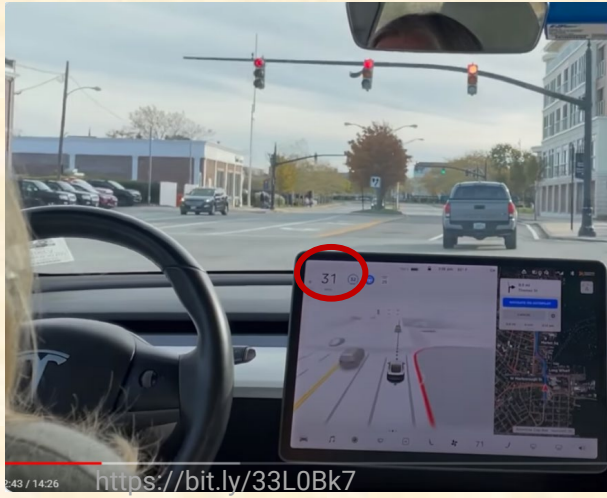
- Key idea: Duty of Care
  - A human driver has Duty of Care to other road users
    - Breach of this duty of care ➔ negligence
  - Must act as a “reasonable person” would act
    - A theoretical competent, unimpaired person, according to a jury
    - Per incident ➔ statistical safety does not avoid negligence

© 2024 Philip Koopman

Duty of Care for Accountability

Carnegie Mellon University

- Legal fiction of a "computer driver"
  - Sustained automated steering of vehicle
  - Manufacturer is responsible
- Transfer of duty of care is key
  - Computer driver has it while steering
  - Can transfer duty of care back to human
    - With sufficient notice



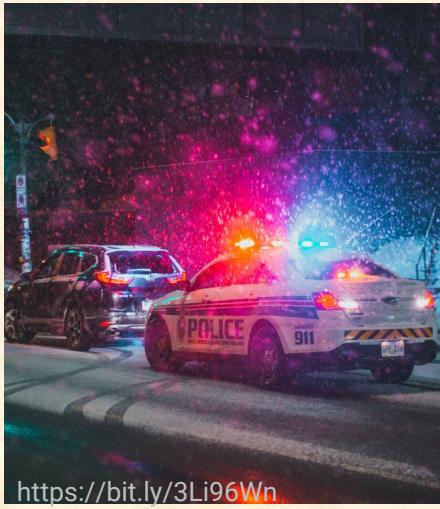
- Computer driver held to same standard as human driver
  - Would a human driver have been negligent?
    - Loss resulting from traffic law violation is negligence per se
  - Statistical safety doesn't avoid negligence (no "free hits")

© 2024 Philip Koopman

Implications of Defining a Computer Driver

Carnegie Mellon University

- Most crashes can be handled by tort law
  - Computer Driver that runs a red light …
    … held to same rules as if a Human Driver
    - Do we really need source code analysis for this?
  - Avoids overwhelming courts with product liability
    - Straightforward fix without rewriting existing law
  - Analogous to "electronic signatures" ➔ signatures
- Financial pressure for safe driving behavior
  - Same rules for Computer & Human Driver behavior
  - Manufacturer bears costs from any unsafe driving
  - Need more for acceptable safety at scale! *But this is a start.*



https://bit.ly/3Li96Wn

© 2024 Philip Koopman

Alternative to SAE Levels for Regulation

Carnegie Mellon University

- Conventional: Human Driver steers
  - Human Driver responsible

- Fully Autonomous: Computer Driver steers
  - Manufacturer is responsible for Computer Driver

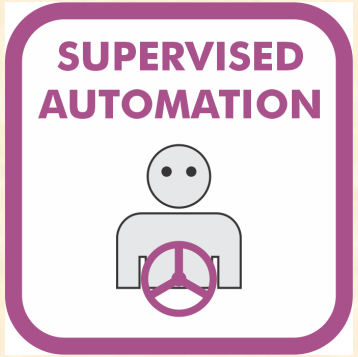
- Testing: Development, Beta, Pre-production
  - Manufacturer is responsible for safe test plan, qualification and performance of test drivers

© 2024 Philip Koopman

The Awkward Middle

Carnegie Mellon University

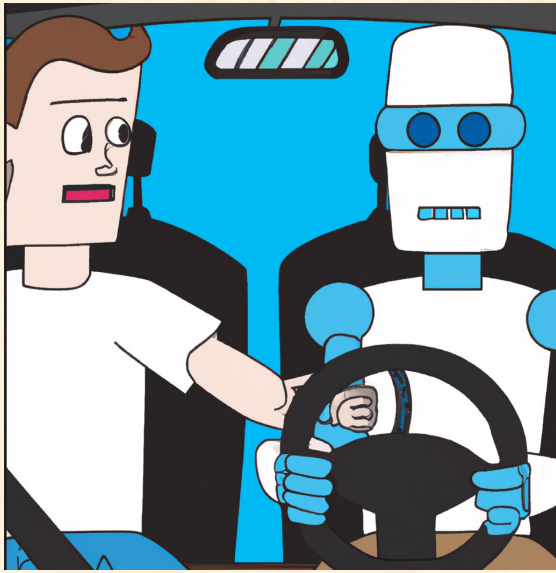
- Unify SAE Levels 2/3 into single regulatory bin
  - Computer steers + other control; human supervises
- Activated computer driver accepts duty of care
  - Human role determined by operational concept
- Computer driver can relinquish duty of care:
  1. Due to driver monitor violation
  2. Due to exiting Operational Design Domain
  - But only after 10 second minimum safe harbor for human driver
    - Best effort fault mitigation after 10 second timer
    - Longer safe harbor if jury says this is reasonable for situation



© 2024 Philip Koopman

Providing A Safety Guardrail

- Automated steering is the key safety attribute
- Net risk metrics are insufficient
  - Safer than human is a long term goal
  - Will take years for equipment regulations
  - What about risk redistribution & inequities?
  - Solutions needed, *but will take time*
- Computer Driver concept
  - Compatible with what many companies are selling
  - Imposes same requirements we already use for human drivers
  - Holds companies accountable for cost of mishaps



[Dall-E]

Tort Law Could Help Support Safety – Via Computer Driver Concept

Essential Vehicle Automation Safety

Carnegie Mellon University

1. Safe as a human driver on average
   - Perhaps 100M miles/fatal crash (better for good drivers)
2. Avoiding risk transfer onto vulnerable populations
   - Pedestrian harm should not increase even if net harm is reduced
3. Avoid negligent computer driving
   - Running red lights and stop signs is not OK
4. Conform to industry safety standards
   - Uncrewed operation = deployment
5. Address other ethical & equity concerns
   - Limited local authority; manufacturer accountability for harm

© 2024 Philip Koopman

Resources

Carnegie Mellon University

- Video lecture series on autonomous vehicle safety:
  - Keynote talks: https://users.ece.cmu.edu/~koopman/lectures/index.html#talks
  - Mini-course: https://users.ece.cmu.edu/~koopman/lectures/index.html#av
- "Safe Enough" book & talk video:
  - https://safeautonomy.blogspot.com/2022/09/book-how-safe-is-safe-enough-measuring.html
- UL 4600 AV safety standard book & talk video:
  - https://safeautonomy.blogspot.com/2022/11/blog-post.html
- Liability-based proposal for state AV regulation & podcast
  - https://safeautonomy.blogspot.com/2023/05/a-liability-approach-for-automated.html
- US Congressional House E&C testimony:
  - https://safeautonomy.blogspot.com/2023/07/av-safety-claims-and-more-on-my.html

© 2024 Philip Koopman