



Breaking the Tyranny of Net Risk Metrics for Automated Vehicle Safety

October 2023

www.Koopman.us

Carnegie Mellon University

PHILIP KOOPMAN

HOW SAFE IS SAFE ENOUGH?

Measuring and Predicting Autonomous Vehicle Safety



What Do We Mean By Safe?

- Rhetoric:
 - “Safety is our #1 priority”
 - “Robotaxis won’t make stupid driving mistakes”
- Acceptable net risk:
 - Positive Risk Balance
 - Risk is managed via insurance
- Requirements beyond net risk:
 - Avoid risk inequities
 - Avoid negligent driving behavior
 - Expectation of safety via engineering rigor



[Dall-e]

“Safety Is Our #1 Priority”



Because
Safety is
Urgent™

Autonomous Driving
Technology Can Save
Lives and Improve
Mobility

<https://waymo.com/safety/>

cruise

Safety first,
always

<https://getcruise.com/safety/>



Safety Drives Us

Motional is developing safe
autonomous vehicles.

<https://motional.com/safety-philosophy>

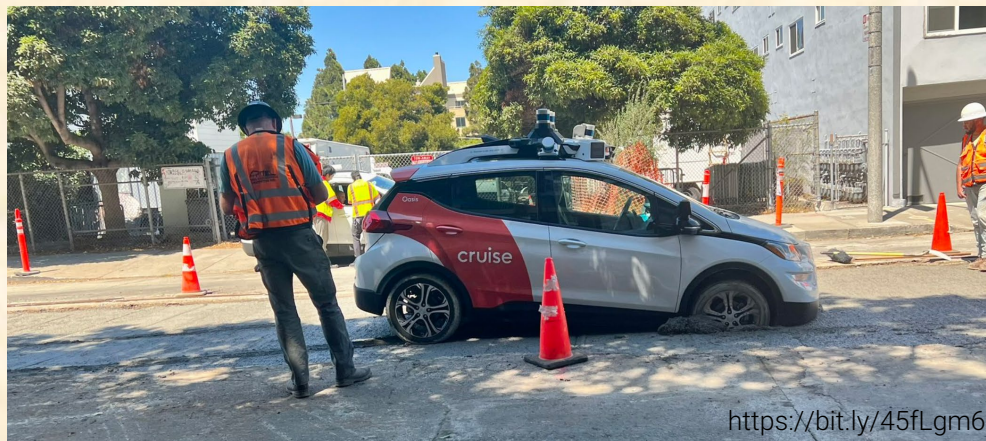
zoox

A new bar for safety

Safety isn't just part of what we do. It's why we're here.

<https://zoox.com/safety/>

Do Robotaxis Make Stupid Mistakes?



Positive Risk Balance (PRB)

- PRB: No worse than a human driver
- Human driver baseline for comparison??
 - Which driver (age, training, impairment)?
 - Where (region, road type, road condition)?
 - When (weather, lighting, congestion)?
 - Which vehicle (new with AEB, or old junker)?
- **Difficult to confirm at deployment time**
 - Need 100M+ miles to determine outcomes
 - Simulations have limited ability to predict edge case outcomes
 - Each crash during data gathering presents bad optics for industry



[Dall-e]

Insuring Risk

- “We are safe because we bought insurance”
 - Insurance companies struggling to evaluate AV risk
- Affordable risk might exceed everyday safety
 - Commercial space launch insurance
 - Life insurance for combat military personnel
 - Insurance is about pricing risk, not ensuring safety
- Property damage can outweigh cost of harm
 - Motorcycle insurance cheap – less property damage
- Affordable Insurance \neq Acceptable Safety



<https://bit.ly/46umY8J>

Limits To Statistical Safety

■ Redistribution of harm

- What if more pedestrians, cyclists die?
- What if more mishaps happen in historically disadvantaged areas?

■ Negative risk externalities

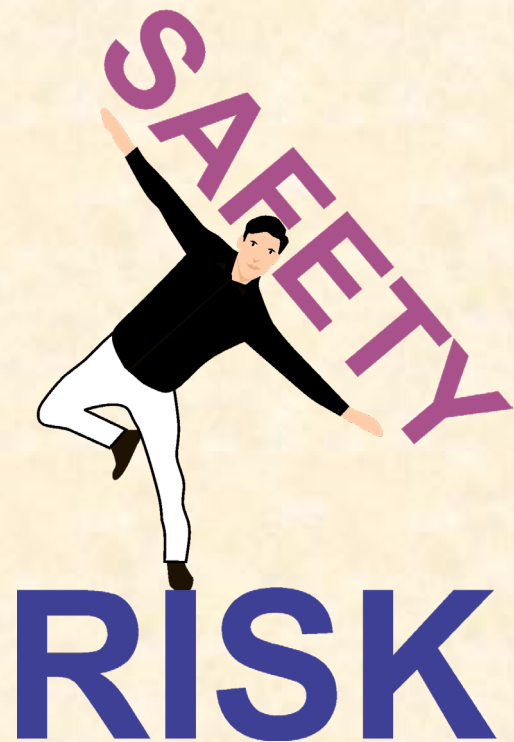
- Blocking fire trucks, ambulances

■ Known significant risks not mitigated

- 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*



■ Civil Tort Law

- Compensate a claimant who has suffered loss ... proximately caused by ... the **negligence** of another party.

■ Key idea: Duty of Care

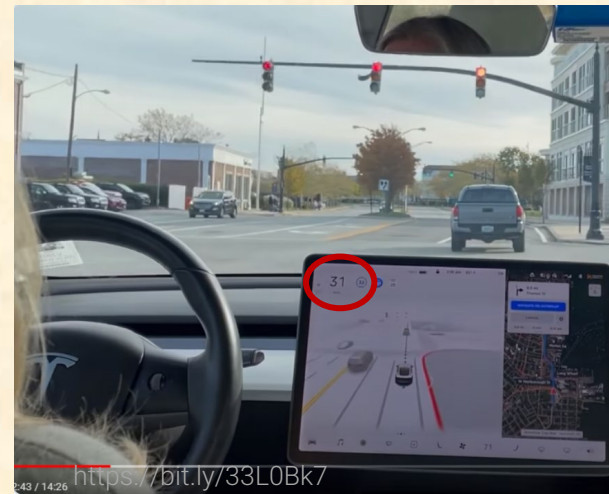
- A human driver has **Duty of Care** to other road users
 - Breach of this duty of care → negligence
- Must act as a “reasonable person” would act
 - A theoretical competent, unimpaired person, according to a jury
 - Per incident → statistical safety does not avoid negligence



<https://bit.ly/3KO9PPe>

Negligence → Accountability

- 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 (10 seconds or more)
- 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 does not avoid negligence

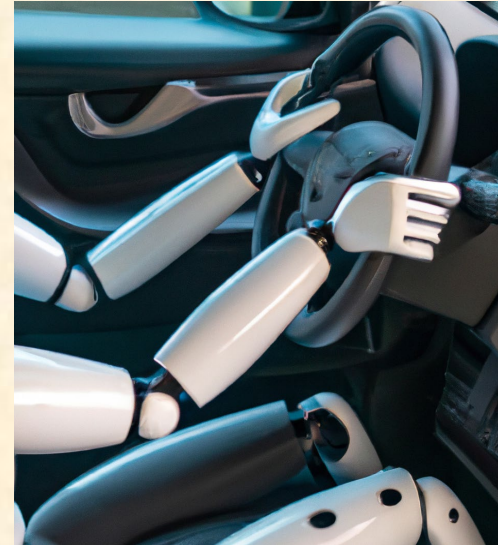


- **Need process-based assurance**
 - Pure testing is impractical; lagging metric
 - Simulations might have defects
 - We get safety via engineering rigor
- **Safety standards + good engineering**
 - Sets prior expectation of acceptable safety
 - Shows good faith efforts for safety
 - Might be strongest credible argument for deployment
 - ISO 26262, ISO 21448, UL 4600, ISO 21434, etc.



Summary: Safe Enough AV Deployment

- Rhetoric is just talk
 - Need a strong safety culture
- Net risk metrics
 - Risk management is just a start
 - Safer than human is a long term goal
- Beyond statistical safety
 - Avoid risk inequities
 - Avoid negligent driving behavior
 - Avoid regulatory-based recalls
 - Use engineering rigor & oversight to set expectation of safety



[Dall-e]

- Liability-based proposal for AV regulation & podcast
 - <https://safeautonomy.blogspot.com/2023/05/a-liability-approach-for-automated.html>
- Video lecture series on autonomous vehicle safety:
 - Keynote AV Safety overview video : https://youtu.be/oE_2rBxNrFc
 - 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>

QUESTIONS?