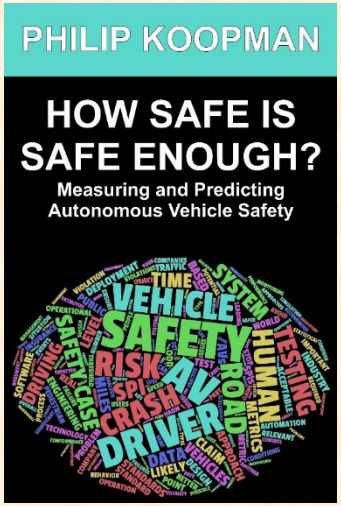
Philip Koopman



Defining a Computer Driver for Automated Vehicle Accountability

October 2023

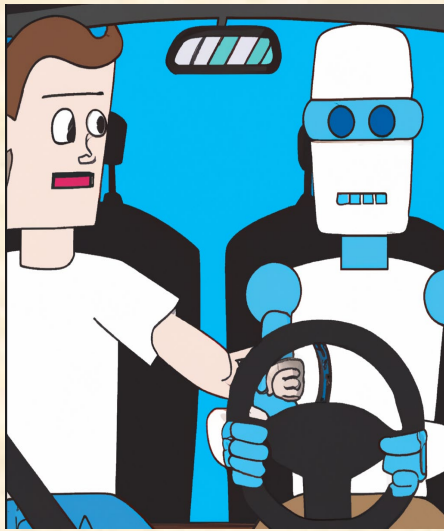
www.Koopman.us

Carnegie Mellon University

Overview

Carnegie Mellon University

- Common solutions to ensure AV safety insufficient
  - "Safer than human" will take years to prove
  - Regulations will take many years
  - Product liability alone is not viable
- Solution: define a computer driver
  - Duty of care equivalent to human driver
    - Manufacturer responsible for negligent behavior
  - Transfer of duty of care must be clear



[Dall-E]

- Provides more accountability for automated vehicle safety

Product Liability Is Not Enough

Carnegie Mellon University

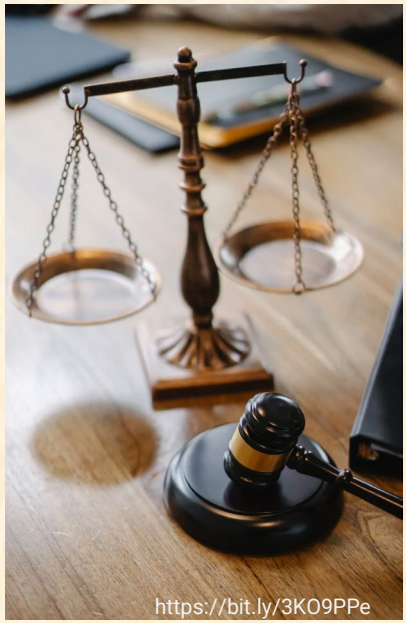
- Manufacturers are pushing for only product liability
  - Manufacturing defect, design defect, etc.
  - Product proven to present undue risk
- Difficult and expensive to prove
  - Source code analysis expensive + painful
  - Class action requires commonality
    - With weekly neural network updates?
  - 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.

Tort Law for Engineers

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

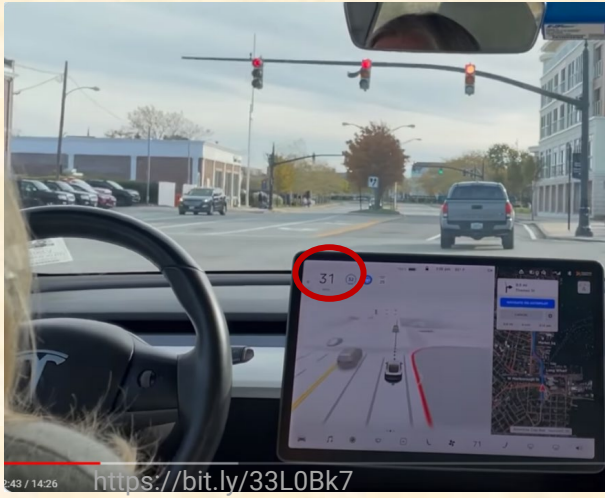


- Key idea: Duty of Care
  - A human driver has Duty of Care to other road users
    - Breach of this duty of care ➔ negligence
  - Must act as a “reasonable person” would act
    - A theoretical competent, unimpaired person, according to a jury
    - Per incident ➔ statistical safety does not avoid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

Duty of Care for Accountability

Carnegie Mellon University

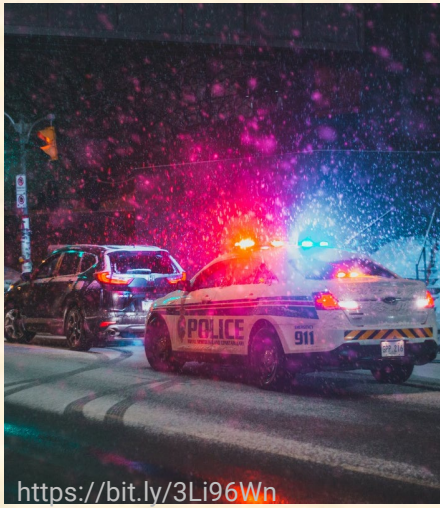
- Legal fiction of a "computer driver"
  - Sustained automated steering of vehicle
  - Manufacturer is responsible
- Transfer of duty of care is key
  - Computer driver has it while steering
  - Can transfer duty of care back to human
    - With sufficient notice
- Computer driver held to same standard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



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.*

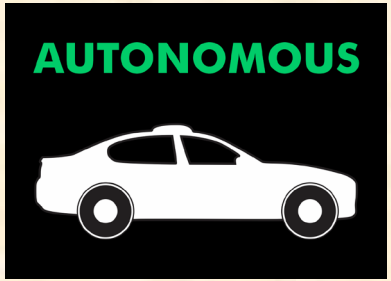


https://bit.ly/3Li96Wn

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

The Awkward Middle: Supervisory Mode

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

Urgency of Defining a Computer Driver

Carnegie Mellon University

- Automated steering is the key safety threshold
- 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*

[Dall-e]

- 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