University of Toronto
Department of Computer Science

Autonomous Vehicle Safety:
An Interdisciplinary Challenge

Presenter: Yasaman Rohanifar
MSc Student, Department of Computer Science
University of Toronto

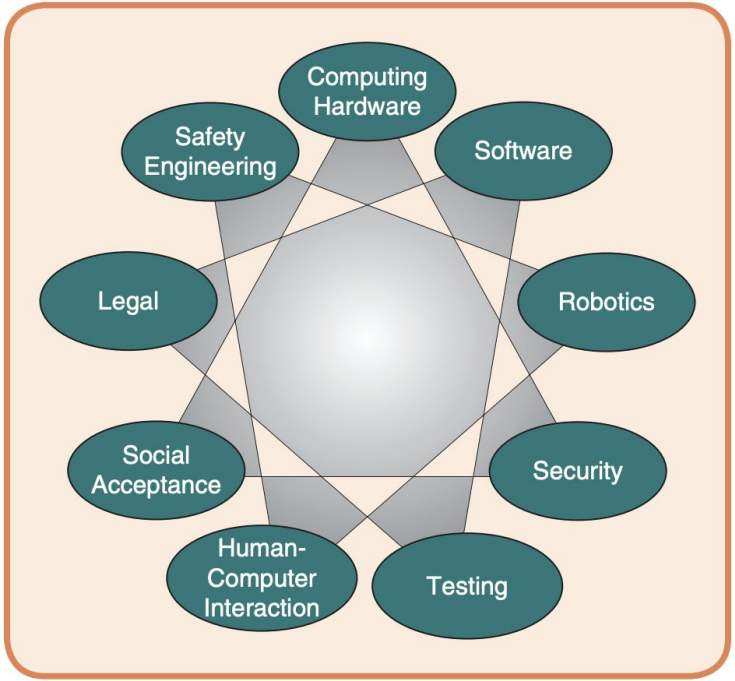
February 25, 2019

About this paper

- IEEE Intelligent Transportation Systems Magazine
- Spring 2017

About the Authors:

- Philip Koopman - Carnegie Mellon University
- Michael Wagner - Edge Case Research LLC



Source: paper

Fish Eye View of the paper

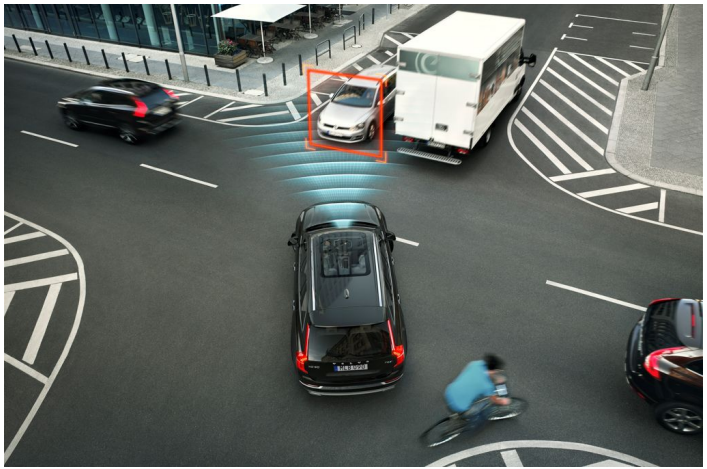
- Call for a **multidisciplinary approach** across all levels of functions and hierarchy to ensure safety for fully autonomous vehicles.
- Validation of inductive learning against new environment inputs as an open technical problem.
- An approach to an end-to-end design and development process that addresses the safety concerns of different specialties.

Source: http://www.uwphotographyguide.com/images/car-photo.JPG

Introduction

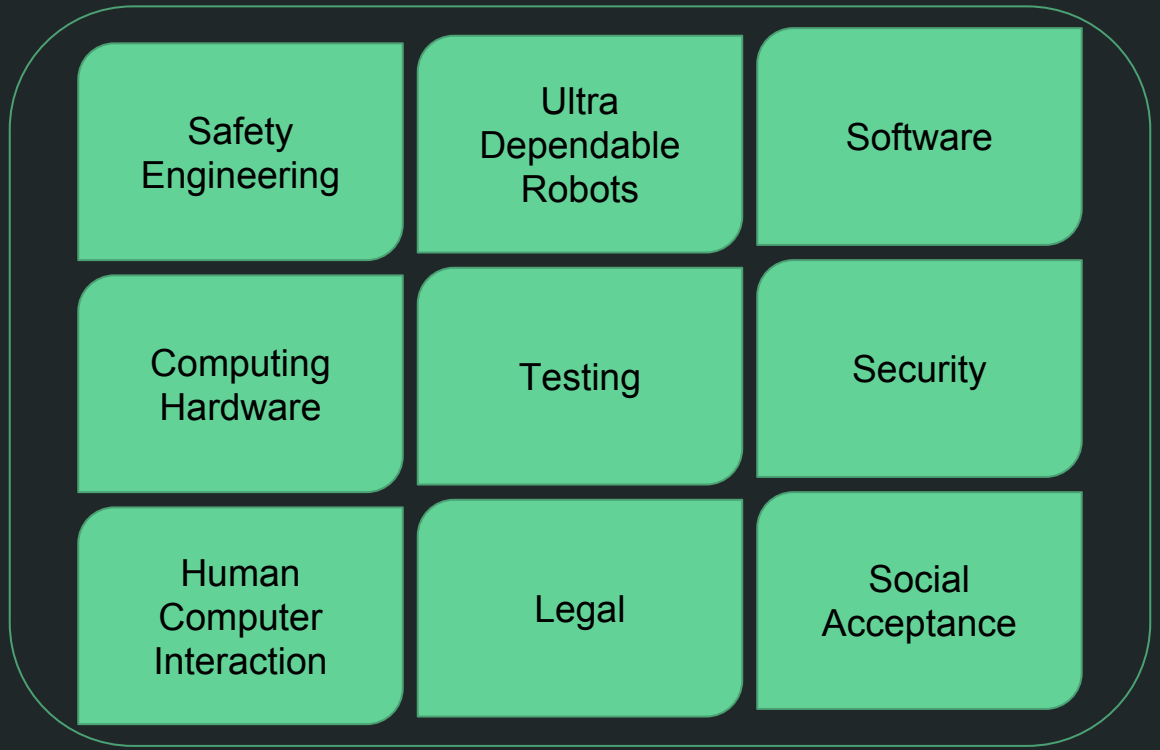
- Dream of fully Autonomous cars
  - Relieving people from the stress of driving
  - Reduction in driving fatalities
- Autonomous Vehicles won't be perfect!
  - When will be able to safely deploy a fleet of fully autonomous vehicles?
- Nothing is simple with autonomous vehicles, even the definition of safety!
  - Implementation of vehicle-level behaviors? Dealing with hazards?
  - Failover mission planning?
  - A method to validate inductive-based learnings?

Source: https://s4745.pcdn.co/wp-content/uploads/2018/10/womaninautomatedcar.jpg

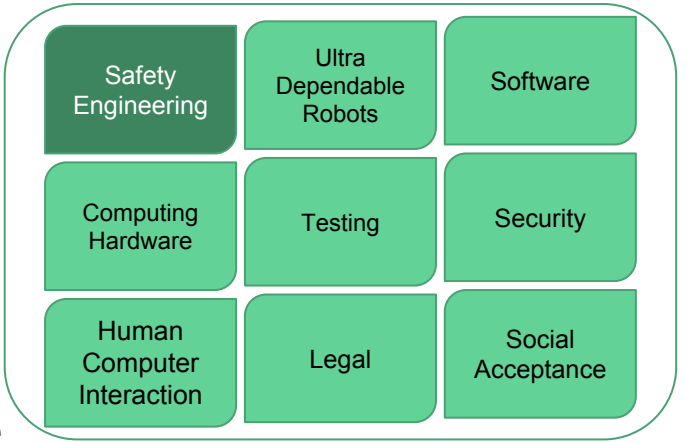


Source: https://www.extremetech.com/wp-content/uploads

Disciplines

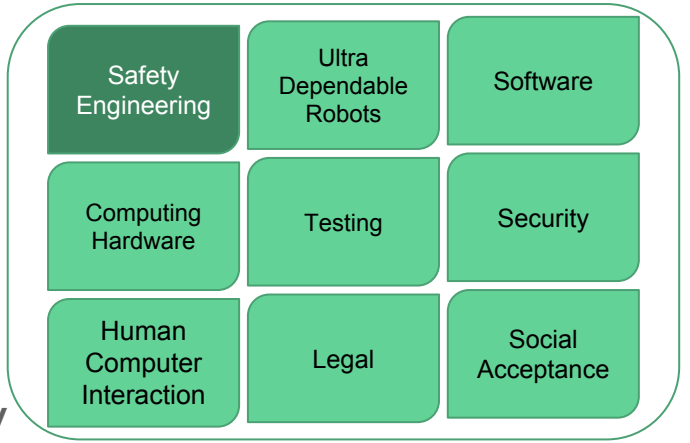


Disciplines: Safety Engineering

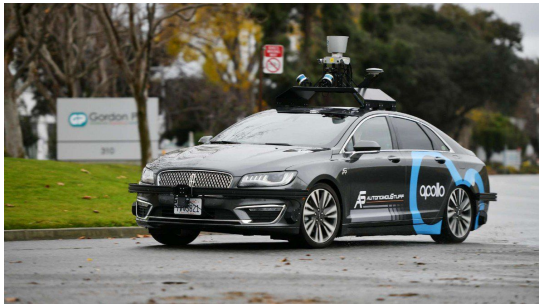


- Challenges of deploying L4 at scale:
  - Managing failures that are infrequent for single vehicles But happen too often when exposure increases!
- Making computer-based automotive systems safe
  - ISO 26262 says a human driver is ultimately responsible for safety…
  - **Solution**: Setting the “controllability” aspect of autonomous systems to zero?
- Unconstrained adaptation such as real time learning of new behaviors result in different behaviors during operation than displayed in testing…
  - Current certification approaches can’t handle these situations
    - The need to consider all system possibilities up front!
  - **Putting limits on adaptation** and **fully explored system design**
  - Formal method approaches ➔ assumptions that might not be provable or testable

Disciplines: Safety Engineering

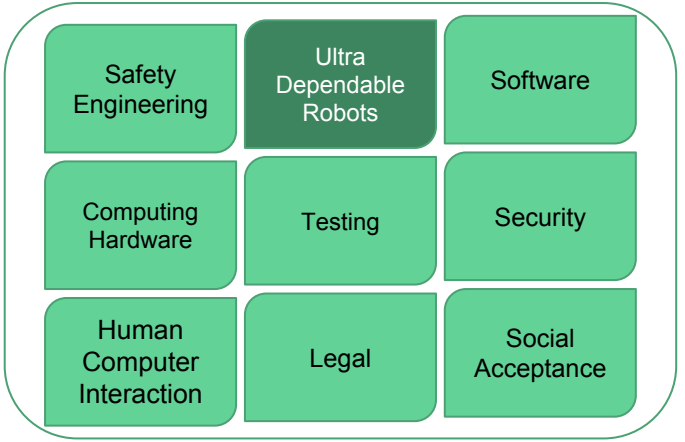


- Creating safety-critical computer-based systems
  - Re-engaging the driver in case of equipment failure
  - Creating **a human safety net** for automation
  - Vehicle must have **a fail-operational autonomy capability**
  - Achieving safe state: cars vs. aircrafts
  - **Strategy**: changing operational modes to short duration “safety mission” in critical situations
    - Designing a smart-enough car that pulls over instead of one that is full autonomous
    - If a safe mission is always available, the primary Autonomy need not be fully operational
    - Relaxing safety requirements on primary autonomy Results in reduction in costs and complexity



Source: https://cdn.technologyreview.com/i/images/apollocar.jpg

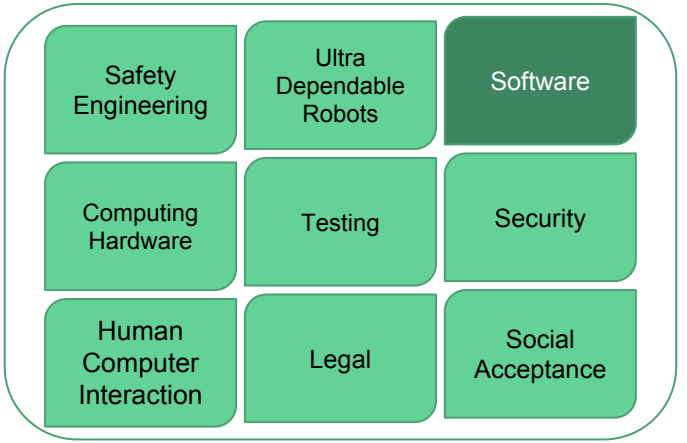
Disciplines: Ultra-dependable Robots



- Goal of making fully autonomous cars as safe per Operating hour as aircraft
  - Safety level ~ 1 billion operating hours/catastrophic event
  - Ultra-dependability!
- Challenges of ultra-dependability:
  - Improving the robustness of the system for messy environments (e.g. debris, clutter, ...)
  - A system that self-monitors its confidence in operation
    - High false-negatives ➔ unintentional unsafe behaviors
    - High false-positives ➔ leaving too many cars stranded
  - Validation of inductive reasoning used in ML techniques
    - Difficult to reason about the correctness of the ML systems Behavior in the face of novel data.

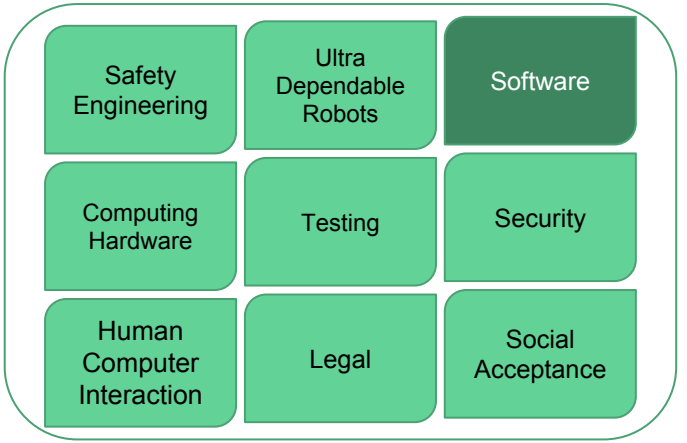
Source: https://cdn.cnn.com/cnnnext/dam/

Disciplines: Software



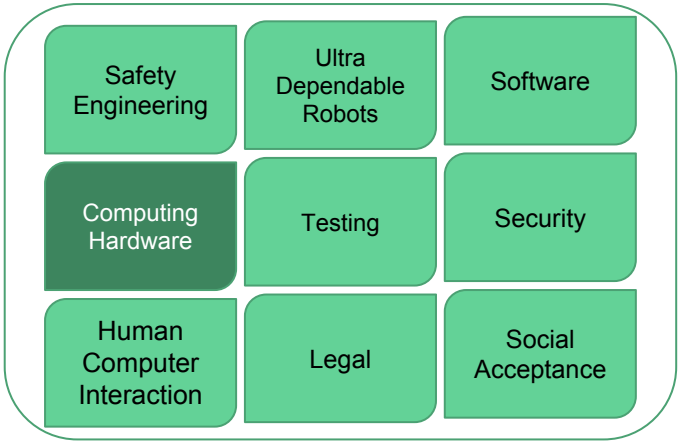
- Current software safety approaches: V Process
  - Assumes HQ requirement are refined into implementation
- Challenges with adaptive and ML systems
  - Can they confront the messiness of the real world?
  - Are ML validation sets comprehensive enough?
  - Knowing that the training and validation sets are good enough is not easy! Edge cases?
- Even if the validation sets are extremely comprehensive…
  - What about the “Unknown unknowns”?
  - Some operating scenarios may seem ordinary to a person and not included in the test data set

Disciplines: Software



- Basing a safety argument on the sufficiency of training and validation data potentially makes the system that collects this data safety critical!
- **Potential solution:** defining “safe” operation in an independent and safe way.

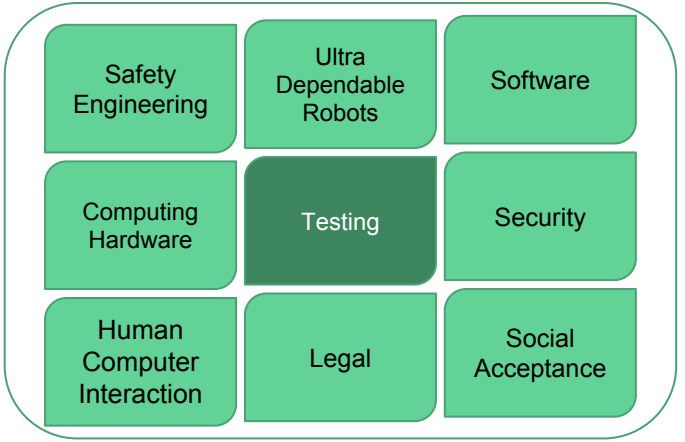
Disciplines: Computing Hardware



- Ultra low cost hardware with safe failure behavior
- Combined hw/sw architecture that employs redundancy policy
- Latent fault detection
  - Even 1-2% of gaps in self-diagnosis has dramatic implications for achieving reliability
  - Undiagnosed failures can accumulate for the entire working life of the vehicle
  - Probability of experiencing multiple independent undiagnosable failures during vehicles lifetime will be higher than one single driving mission
- Need for chips that can be self-tested before each driving cycle with high testing coverage

Disciplines: Testing

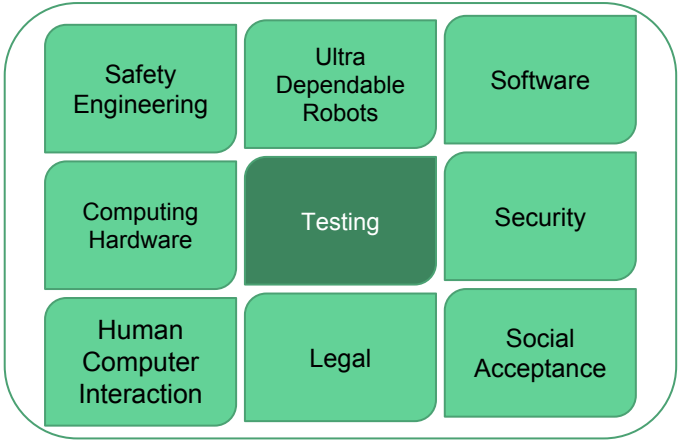
- Traditional pre-computer car testing
- Recent AV on-road testing
- Testing only approach is insufficient
  - What about full-fleet deployment?
- V-Model
  - Comparing a defined design document against a system
- Probabilistic Systems
  - Behavior of the system is expected to be different in each run
  - Small changes in initial conditions result in large changes
  - The testing oracle should support abstract results not specific ones!



Source: https://s.hswstatic.com/gif/car-testing1.jpg

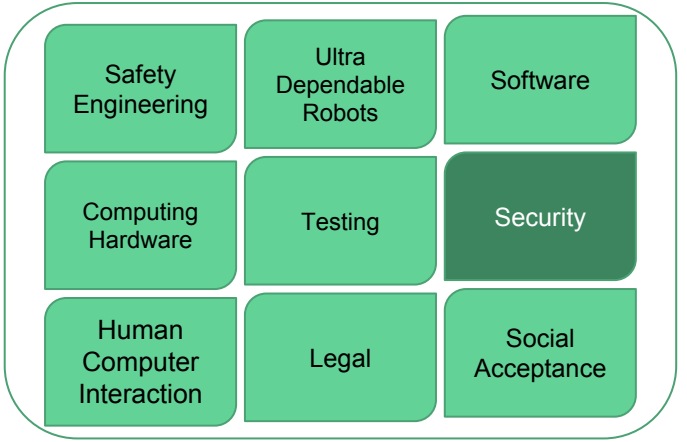
Source: https://s.abcnews.com/images/GMA/

Disciplines: Testing

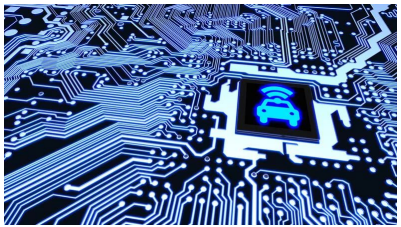


- Challenges of inductive-learning systems
  - No design and starting point for testing oracle
  - Sets of training data and validation data
- Even comprehensive data is not enough!
  - Characteristics of the training data or coincidental correlations?
  - Measurements on the test data may be different than actual data in the future!
- Ultra-dependability for machine learning algorithms is still uncertain!
- Fault injection and failure management
  - Is the vehicle controllability fully responsible for the autonomous system?
  - How would the vehicle deal with a tire blow out or autonomy algorithm failure?

Disciplines: Security

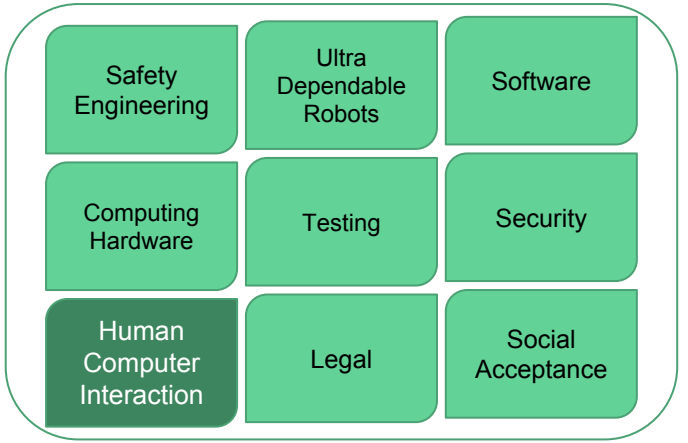


- Measures for:
  - Specific vehicles
  - System-level attacks and failures
- Can't trust the security of other vehicles or even roadside infrastructure!
  - What if in v2v encrypted communication a vehicle provides maliciously incorrect information?
  - What if someone physically breaks into roadside infrastructure and reprograms it?
  - What if the power is killed by a malicious person?
- It's naive to think that standalone vehicles will be able to:
  - Realize they are being fed incorrect or malicious information
  - Detect an attack
  - Perform safing mission under attack



Source: https://i0.wp.com/readyspace.com/

Disciplines: HCI

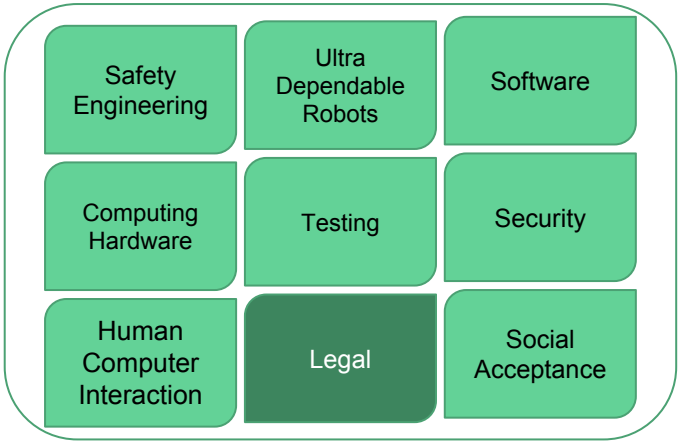


- Even full-autonomous vehicles must make sure
  - Occupants feel safe
  - Build customer trust (vital for technology adoption)
  - Anticipate behaviors of other vehicles
- How AVs interact with human drivers of other vehicles?
  - Even with full AVs, there are bicycles, scooters, motors, etc.
  - Pedestrians (especially ill-behaved/children/pranksters)
  - Public pressure to spread autonomy!
- AVs actions should be easily perceivable by human
  - Human-Vehicle interaction safety
  - Testing coverage
  - Design comprehension

Source: https://www.researchgate.net/profile/Amir_Rasouli4

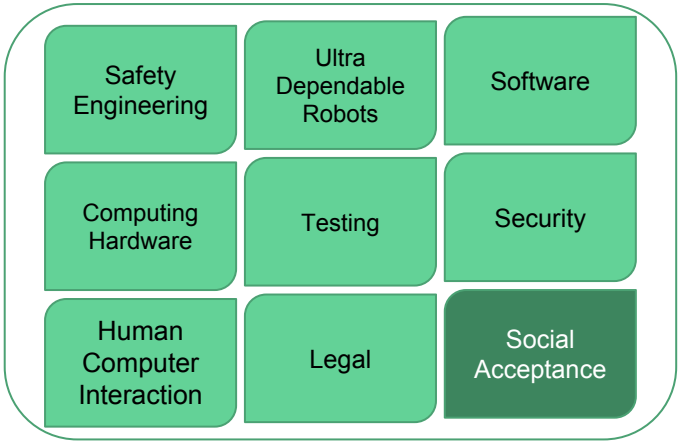
Source: https://www.autocar.co.uk/sites

Disciplines: Legal



- Open legal issues of liability
- Vehicle logs as sources of information
  - Can we trust the data from a malfunctioned vehicle?
  - Independent data recording system for mishap forensics
- Who is responsible for proper vehicle operation?
  - Vehicle manufacturer that trusted a faulty third-party data set?
  - The mechanic who installed an incompatible sensor?
  - Operating system vendor who didn't deploy a security patch fast enough?
- Do we have the adequate technological foundation to build upon?

Disciplines: Social Acceptance



- AVs will be safer drivers than people
  - Unrealistic!
  - Avoiding collisions that are physically impossible
- What is the standard for autonomous safety?
  - Better than excellent? Typical driver?
  - How is a “typical driver” characterized?
  - Tricky situations
    - Ordinary human drivers have a good chance of avoiding mishap but the autonomous vehicle crashes
- Establishing statistical basis for insurance purposes
  - Suitable application of monetary reserves - **Vehicle vendors acting as reinsurers!**

Summary

P1	Call for safety certification strategy of some sort for fully autonomous vehicles (cross-disciplinary approach),
P2	Edge cases, subtle trade offs, cross-coupling trade offs between areas, updating practices and validation processes to address safety concerns and lead to an end-to-end design.
P3	Challenges with validating ML based systems to ultra-dependable levels required for autonomous vehicles fleet

Discussion Points/Questions

Q1	Are current standards and safety measures (such as ISO 26262) sufficient for deployment of self-adaptive autonomous vehicles?
Q2	Does safety even makes sense after setting the “controllability” aspect of autonomous systems to zero in ISO 26262?

Thank you for your attention!

Questions?