A Safety Framework for Shared Human/Computer Driving Responsibility

June 30, 2023

www.Koopman.us
Overview

“Computer Driver” as a concept
- Same duty of care as a human driver
- Perform as a “reasonable driver”

What about shared responsibility?
- Effective driver monitoring
- Reasonable responsibility transfer process

State liability laws play a key role
- Buys time to sort out equipment regulation
- Can work with a non-statistical definition of “safe enough”
Key Approach: Computer Driver

- Need more than statistical approach when computer drives
  - Challenges to predicting initial safety outcomes
  - Defective behaviors masked by net safety improvements
  - Risk redistribution to vulnerable populations

- Computer Driver should have a duty of care
  - Obligation to be a “reasonable driver”
    - Same criterion as for human driver negligence

- Comparison is “reasonable human driver” ...  
  ... not “average human driver”
  - Manufacturer is responsible party for negligent computer driving
Three “Pure” Operational Modes

- 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

- Human Supervises automated Control of steering
  - Computer Driver has sustained control of steering
  - Prone to Human Driver automation complacency

- This mode includes:
  - Driver told secondary tasks forbidden/acceptable
  - Hands on/off wheel
  - Eyes on/off road

- Unify SAE Levels 2-3 into single, flexible regulatory approach
Moral Crumple Zone: [Elish 2019]
- Blaming nearest convenient human for an automation failure

Ineffective ways to improve safety:
- Blaming humans for exhibiting human error
- Blaming victims
- Liability immunity for manufacturers

Backup Driver Of Autonomous Uber SUV Charged With Negligent Homicide In Arizona

Tesla driver charged with manslaughter in deadly Autopilot crash raises new legal questions about automated driving tech
A Tesla Model S driver accused of crashing his car while Autopilot was activated had run a red light and slammed into a Honda Civic, killing its occupants.
Rule #1: Driver Monitoring Rule

- Manufacturer responsible for distracted Human Driver crash unless:
  - Effective distracted driver alert activated, AND
  - Alert lasts at least 10 seconds before crash, AND
  - Computer Driver ensures safety for those at least 10 seconds.

- Exception:
  - Malicious defeat of driver monitor

Rule #2: Driver Intervention Rule

- Manufacturer responsible for Human Driver failure to intervene unless:
  - Undue risk of mishap readily apparent with enforced level of attentiveness, AND
  - Human Driver has adequate opportunity to intervene – Safe harbor for first 10 seconds

- Computer Driver can demand that Human Driver intervene – but must follow this rule

https://bit.ly/33L0Bk7
Implications: “Readily Apparent”

- Must be obvious deviation from safe driving
  - Computer Driver deviates from its customary behavior
  - Conventional driver would recognize a danger
    - Given only amount of attention that is enforced

- Alarms can make issues readily apparent:
  - ODD departures
  - Equipment failures

- Operational concept affects this
  - Eyes-on-road makes road hazards more apparent
  - Eyes-off-road concepts make hazards less apparent

Human driver readiness
- Attention and tasking status both matter

Time to react
- Enough time appropriate to circumstances
  - Time to recognize Computer Driver acting unsafely
  - Time to switch tasks
    - What if watching a movie?
    - What if hands full?
  - Complexity of road situation, severity of failure, etc.
- Competent (not expert) driver can reasonably intervene successfully
- Computer Driver ensures safety during reaction time
Summary: Driving Safety Responsibility

- **Autonomous mode**
  - Manufacturer – not owner, not the computer itself

- **Testing mode**
  - Test driver might contribute, but not a scapegoat

- **Supervisory mode**
  - Manufacturer except:
    - Rule 1: Human Driver ignores effective driver monitor
    - Rule 2: Human Driver had a fair chance to intervene
  - Manufacturer must respect inherent human limits
What Happens Next?

- Implementation: State Liability Laws
  - Sets a well-defined playing field for liability
  - Based on “reasonable” driver behavior
    - Uses same legal rules applied to human drivers
    - Source code analysis not required

- Technical implications
  - Indirectly regulates driver monitoring effectiveness
    - Can only take credit for driver attention that can be monitored
    - Monitoring sophistication higher for aggressive operational modes
  - Indirectly affects viable concepts of operation
    - Disincentivizes some moral crumple zone strategies

Resources

- Liability-based proposal for AV regulation & podcast
  - [https://safeautonomy.blogspot.com/2023/05/a-liability-approach-for-automated.html](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](https://youtu.be/oE_2rBxNrfc)
  - Mini-course: [https://users.ece.cmu.edu/~koopman/lectures/index.html#av](https://users.ece.cmu.edu/~koopman/lectures/index.html#av)

- “Safe Enough” book & talk video:

- UL 4600 book & talk video: