A Reasonable Driver Standard for Automated Vehicle Safety

Philip Koopman, William H. Widen

Carnegie Mellon University, Pittsburgh PA, USA
University of Miami – School of Law, Miami FL, USA
koopman@cmu.edu, wwiden@law.miami.edu

Abstract. Current "safe enough" Autonomous Vehicle (AV) metrics focus on overall safety outcomes such as net losses across a deployed vehicle fleet using driving automation, compared to net losses assuming outcomes produced by human driven vehicles. While such metrics can provide an important report card to measure the long-term success of the social choice to deploy driving automation systems, they provide weak support for near-term deployment decisions based on safety considerations. Potential risk redistribution onto vulnerable populations remains problematic, even if net societal harm is reduced to create a positive risk balance. We propose a baseline comparison of the outcome expected in a crash scenario from an attentive and unimpaired "reasonable human driver," applied on a case-by-case basis, to each actual loss event proximately caused by an automated vehicle. If the automated vehicle imitates the risk mitigation behaviors of the hypothetical reasonable human driver, no liability attaches for AV performance. Liability attaches if AV performance did not measure up to the expected human driver risk mitigation performance expected by law. This approach recognizes the importance of tort law to incentivize developers to continually work on minimizing driving negligence by computer drivers, providing a way to close gaps left by purely statistical approaches.

Keywords: Automated vehicles, safety, liability, regulation

1 Introduction

Discussions about autonomous vehicle (AV) safety requirements historically have two major themes: claiming that human drivers are bad, and a statistical/utilitarian argument of reduced net harm. However, neither ensures acceptable safety across relevant stakeholders. In this position paper we argue that a third approach based on a comparison of AV performance to a "reasonable driver" is required. While one might parlay a "Positive Liability Balance" into an AV deployment standard based on lower net liability than human drivers, we argue that the statistical approach must be combined with practical accountability for negligent AV driving behavior assessed crash-by-crash.

Industry promotional narratives over-emphasize the first theme of human driver frailties, while promising that computer drivers will necessarily improve safety because computers do not have problematic human behaviors like drunk driving [1]. This narrative ignores potential AV crashes caused by software defects.

The second theme uses an engineering-based narrative that makes utilitarian arguments claiming improved safety from AV deployment—the Positive Risk Balance (PRB) label favored in Europe and, the Absence of Unreasonable Risk (AUR) label more favored in the US [2]. Reducing overall harm is a worthy goal, but the devil is in the details for PRB, which posits a comparison of AV safety outcomes to some notional average human driver safety outcome. To the degree that the AUR label uses a metric of "better than a human driver," both labels end up in the same place.

Whether PRB losses involving automation are comparable to losses generated by human drivers depends on operating conditions, passive safety features, active safety features, risk due to geographic location, risk due to driver age, and so on. Accurately predicting risk outcomes for novel technology before deployment is a significant challenge which raises serious ethical concerns because the technology may cause harm not only to vehicle occupants, but also to other road users who did not volunteer for this grand AV experiment by purchasing an AV or electing to ride in a robotaxi.

Fully autonomous AV proponents complain that, though deployment will (they presume) save many lives, their crusade to improve road safety will be impeded both by regulation and imposition of liability on their technology for every crash [3].

Previous approaches to managing risk and regulating technology propose pursing insurance-based solutions, enhancing information sharing, technology regulation, and making significant changes to product liability approaches [4]. While these approaches can all provide value in the long term, we see a need for a minimalist intervention that can substantially improve the safety situation right away, thereby providing time for realignment of industry business and regulatory models as technology matures.

We argue that blaming automated vehicle technology for *each crash a competent human driver might reasonably have avoided* – regardless of net safety outcomes – is a safety promoting feature, not a public policy mistake. In fact, it is the most efficient and effective strategy to resolve many liability claims that will follow deployment, while at the same time providing a non-regulatory approach to motivate the industry to make responsible deployment decisions for this still-immature technology.

2 Limits To Statistical Safety Approaches

A rhetorical approach claiming that AVs will be safe because "computers don't drink and drive" has convinced some state legislators to allow AVs on the road and to gather limited statistical data in the process. However, it will be many years before the hundreds of millions of miles required for statistical significance will accumulate. Even then, statistical approaches will have issues beyond the challenges of performing a like-for-like comparison to an appropriate human driver baseline [5].

Even if net statistical parity were achieved overall between AVs and humans, the shift to AVs might be unacceptable due to risk redistribution. This is especially true if vulnerable and disadvantaged populations (likely vulnerable road users) have increased risk exposure, while safety benefits accrue to economically advantaged AV users.

Consider a thought experiment in which all cars instantly turn into AVs, decreasing annual US road fatalities from 40,000 to 20,000 per year. We would rightfully celebrate this Positive Risk Balance victory. But some hypothetical outcomes might still prove problematic. (Baseline comparison numbers below are 2020 data from NHTSA [6].)

- All 20,00 fatalities are pedestrians and cyclists. This increase almost triples the 2020 numbers of 6,516 pedestrian and 938 cyclist fatalities (7,454 total).
- The number of fatalities related to school transportation increased to 1000—a nine-times increase over the average of 113 per year.
- Fatalities correlated with people having darker skin color due to a prevalence of light-colored skin in pedestrian training data.
- Technology updates necessitate continuing large-scale road-testing, with residents in testing areas at a significantly increased risk of harm compared to human-driven vehicles, while road users elsewhere enjoy increased safety.
- Every crash was due to a design defect that would likely have been caught by following industry best practices for design quality and validation, but was not.
- Every crash was due to a software defect that was known but not fixed by the manufacturer, due to a rationalization that there was a moral imperative to ship "good enough" software to improve net statistical safety as soon as possible.

We make no judgement as to which of these hypothetical outcomes would be morally justifiable or acceptable by society if they were to occur. However, it is reasonable to anticipate that at least some stakeholders would be unhappy, especially if they were personally affected by increased risk, or even a loss event.

Harm is especially problematic if associated with a risk redistribution, in which some segments of road users suffer increased risk while others enjoy decreased risk. One might devise a fine-grain framework for assessing Positive Risk Balance (a net decrease in risk) to account for risk redistribution. But it is unrealistic to expect this to happen immediately for at least the following reasons: (1) we do not yet know what risk redistribution will actually occur, (2) there will not be enough data to evaluate risk redistribution until after enough harm has been done to build up a statistically valid set of samples. Addressing this topic is essential – but it will take years to do this, while deployment decisions require action now.

3 The Reasonable Driver vs. the Tyranny of Net Risk

One key aspect of statistical statements of safety is that an individual who has been harmed did not benefit from the safety improvement. Even if we solve justice and fairness concerns surrounding statistical risk redistribution, the specifics of each crash still matter for attribution and allocation of liability.

Saying: "We're sorry Widow Smith, but please know that your husband's completely preventable death was part of a statistical 10% improvement in safety. The world is a safer place even as we are meeting our quarterly revenue goals" provides cold comfort to the victim's families. But it would be a shame to lose a potential safety benefit because different people will be harmed by AVs than would be harmed (in greater numbers) by human drivers. (Pareto optimality approaches to resolve this issue are not a viable basis for social decision-making because every choice disadvantages somebody in practice.)

Current AV capabilities do not support a vaccination-type argument to justify ignoring the specifics of each case—a situation in which it is appropriate to ask everyone to take a small risk for the overall public good. Imposing AV risk on the public (including those who do not directly benefit from using AVs at all) might only become compelling when the number of people harmed is reduced by orders-of-magnitude compared to the status quo. Vehicle automation technology is nowhere near such a capability.

Equipment regulation on a statistical basis should continue. Recalls for obvious patterns of harm can be accomplished under existing regulatory frameworks to keep known harms from continuing unabated. Test-based regulations and star-based rating systems can increase pressure on basic safety performance. But we need more.

We propose to add an additional safety approach to risk management and safety standard conformance, which moves from the statistical to the specific: *In every loss event, to avoid liability AV performance must not be negligent as measured by the performance standard we require of human drivers.*

3.1 A Lay Summary of Tort Law

Civil tort law provides a method to compensate a claimant who has suffered loss proximately caused by the negligence of another party. “Negligent driving” refers to unacceptably hazardous behavior by a vehicle (whether human or computer driven). In contrast, “product liability” refers to a manufacturing defect, design defect, or other product characteristic proved to cause unacceptably hazardous behavior.

Pursuing product liability claims typically costs more than pursuing negligence claims because expert engineering evidence proving a design defect or other technical cause of a crash can cost hundreds of thousands of dollars or more. The manufacturer also incurs significant defense costs. Costs and long time-scales render pursuit of product liability claims impractical for a run of the mill traffic accident.

Proving a negligence claim is more straightforward, with the plaintiff claiming that the defendant owed her a duty of care which was breached by failure to act as a reasonable person would have acted, and that this breach caused harm—the legal standard used to determine negligence liability. Breaking traffic laws likely amounts to negligent behavior. If an AV runs a red light, the driver likely is negligent.

Significantly tort law liability can be based solely on negligent behavior without proof of a product defect. Absent a recognized excuse, such as an Act of God, if conduct falls short of the performance we expect of a reasonable person, the defendant has liability. If an AV behaves in a dangerous way, the law should infer AV negligence proximately caused any harm, just as the law infers liability of a human driver. In a rear-end collision, for example, the law tends to infer negligent driving of the rear vehicle.

A crash victim needs a straightforward recovery method in addition to complex product liability claims in AV accident cases, because manufacturers have asymmetric access to data recordings, design expertise, and a war chest to pay lawyers. We must supplement the status quo to better incentivize manufacturers to strive for safety.

3.2 Safety By Comparison To a Reasonable Driver

Using a negligence-based approach to evaluate AV safety changes the baseline for comparison from "anything non-human is better" or "better than a statistically average driver" (Positive Risk Balance) to a human-centric standard. The baseline for liability is the legal construct of a "reasonable" person/driver.

Courts have crafted the concept of "reasonable driver" as a hypothetical ideal person over time for use in resolving human-to-human driver negligence cases. This is not an "average" driver or an expert driver. Rather, this is a driver who can handle situations in a reasonably competent way, without lapses in judgement. For example, if a ball rolls into the street near a school playground, many would agree that a reasonable driver should anticipate a child will soon appear to chase after it and start slowing down before the child appears. This is not a technical standard, but a flexible yet objective standard that has proven its worth for use by courts and lay juries.

Our specific use of this principle is that the actions of an AV should be compared to a hypothetical reasonable human driver with regard to liability for causing harm or failing to mitigate reasonably avoidable harm. If a reasonable human driver would have avoided causing harm, so too should an AV's computer driver. If it were unreasonable to expect a human driver to avoid harm in a particular situation, then the AV driver is blameless – even if a theoretically better design might have avoided the harm. (A claimant still might pursue a product defect claim, but not ordinary negligence.)

On a situation-by-situation basis, an AV should be as safe as a reasonable human driver. Any instance in which the AV displays negligent driving behavior means that it is in practice unsafe, just as a human driver would have been unsafe in that situation.

3.3 Case-By-Case Liability

We propose that tort laws at the US state level should be modified to recognize the concept of negligence for the legal fiction of a computer driver. The idea is that whenever a driving automation system is in control of a vehicle, it should be treated the same as a human driver regarding negligence on a case-by-case basis for every time it might have inflicted harm. The manufacturer should be the responsible party (with the manufacturer perhaps being a system integrator, depending on the specifics) [7].

This approach transforms many highly technical product defect cases into much simpler negligence cases with lower transaction costs to resolve, and more direct access to common-sense outcomes. If an AV makes a mistake and runs a red light when a reasonable human driver in that situation would likely have stopped, the AV manufacturer should be responsible for negligent driving by the computer driver it produced, and incur any burden of proof that exigent circumstances (if any) excused the driving behavior. This is the same burden the law places on a human driver.

AV proponents and engineers in general might argue that exposure to negligence liability creates too great a burden because it might prevent them from making design tradeoffs that reduce overall harm at the cost of increasing redistribution of harm. We argue that such tradeoffs should bear their full social cost. A reasonable design decision should be able to compensate victims while leaving excess manufacturer benefits.

If a human driver with a perfect lifetime driving record crashes while driving the wrong direction on a one-way street, that driver will have negligence liability for losses. A judge might consider the driving record in criminal sentencing, but liability remains. Even the best drivers don't get a "free crash" exclusion from negligence. Computer driver liability should be the same, even if AVs make streets statistically safer.

An added benefit to this approach is that it lowers the bar to reasonable claims by parties harmed by negligent AV behavior. Negligence lawsuits are much less expensive and much more predictable for both parties involved than product liability lawsuits.

Our approach does not require AVs to have perfect, loss-free safety records. Even if they crash, if the technology achieves a "reasonable driver" level of harm mitigation, negligence liability will not attach. Aspirational or factual claims to improved statistical safety should not give AVs a free pass to exhibit driving behavior that would be considered negligent if performed by a human driver.

Requiring behavior that is at least as good as that of a "reasonable driver" raises the question of how to define such behavior in a way that is amendable to deriving engineering requirements. While a reasonable driver is considered well defined from a legal point of view, a precise engineering requirement definition would likely require analysis of case law to account for liability outcomes involving human drivers in a variety of circumstances. While each specific crash might have some aspect of "it depends" involved in assessing liability, there will be common fact patterns that predispose determinations that some computer driver behaviors are reasonable, and some are negligent. For example, a major traffic rule violation such as running a red light in a way that causes a crash is likely to be negligent unless there is significant mitigating circumstance. For borderline cases, the ultimate legal process to determine negligence is via the court system. Engineering audiences might be uncomfortable with a requirement that is not immediately amenable to a prescriptive, deterministic, and complete set of rules. However, they should be able to cope with yet another aspect of vehicle automation which might be addressed via data mining of past court cases to learn what behavior is reasonable.

A technical model might be created that bounds some aspects of reasonable driver behavior. A candidate for some aspects of negligence that might be evaluated is Waymo's "non-impaired human with their eyes on the conflict" (NIEON) model [8].

4 Conclusion

If an AV runs a red traffic light and kills a pedestrian, the victim's family should not suffer through years of uncertainty nor be forced to identify a law firm willing to fund a million-dollar-plus legal effort to hunt for a defect in a machine learning system that caused the vehicle to miss both the red light and the pedestrian in the cross-walk. A much simpler approach presents itself and should be used: simply hold computer drivers to the same standards of negligence as human drivers, with the manufacturer named as the responsible party.

We do not expect AVs to be perfect. But to achieve broadly acceptable safety there needs to be a check and balance mechanism that is both more practical and more immediate than equipment regulation and product defect lawsuits. Tweaking tort law to explicitly support the legal fiction of computer driver negligence provides just such a mechanism. It also introduces a new constraint on AV manufacturers: it is not sufficient to be statistically safe. One must also avoid harm in each situation in which a competent "reasonable driver" would also have avoided harm – the same as human drivers must.

This approach can be extended to encompass not only fully autonomous vehicles, but any vehicle automation technology that is prone to inducing automation complacency or otherwise resulting in a situation in which automation misbehavior is prone to causing mishaps [9]. This includes any vehicle with sustained automated steering, even if a human driver is tasked with supervising automation safety.

This paper concentrates on US tort law. Extending analysis to other legal systems would require future work. However, other legal systems could likely benefit from an alternative to a product defect approach to addressing many loss events that will eventually occur involving automated driving capabilities. Core principles might be introduced to other legal frameworks by defining appropriate equivalents of a duty of care for a computer driver, and a notion of synthetic negligence for computer drivers, with the responsible party being manufacturers.

This same principle of synthetic negligence for computer drivers might be extended to other areas in which automated systems supplant humans who previously owed a duty of care to others. That might help avoid such a duty being swept away or hidden behind the opacity of a product's "artificial intelligence" branding.

References

1. NHTSA, Automated Driving Systems 2.0: A Vision for Safety, DOT HS 812 442, Sept. 2017.
2. F. Favaro, "Exploring the Relationship Between 'Positive Risk Balance' and 'Absence of Unreasonable Risk'", Oct. 20, 2021, https://arxiv.org/abs/2110.10566
3. F. Augustin, "Elon Musk says Tesla doesn't get 'rewarded' for the people its Autopilot technology saves, but instead gets 'blamed' for the people it doesn't," Business Insider, Dec. 19, 2021.
4. B.W. Smith, "Regulation and the Risk of Inaction," In: Maurer, M., Gerdes, J., Lenz, B., Winner, H. (eds) Autonomous Driving. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-48847-8_27
5. P. Koopman, *How Safe Is Safe Enough,* 2023, ISBN: 979-8846251243.
6. NHTSA, Traffic safety Facts, https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813369 accessed June 6, 2023.
7. Widen & Koopman, "Winning the Imitation Game," April 2023, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=442969
8. Waymo, "Benchmarking AV Safety," https://waymo.com/blog/2022/09/benchmarking-av-safety.html accessed June 6, 2023.
9. Widen & Koopman, "The Awkward Middle for Automated Vehicles: Liability Attribution Rules When Humans and Computers Share Driving Responsibilities," May 2023, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4444854