Towards Computational Foreseeability

Sarit Kraus[1], Kayla Boggess[2], Robert Kim[3], Bryan H. Choi[3], Lu Feng[2]

[1]Bar-Ilan University, [2]University of Virginia, [3]Ohio State University
sarit@cs.biu.ac.il, {kjb5we, lu.feng}@virginia.edu, {kim.9566, choi.1399}@osu.edu

Abstract

This paper addresses the challenges of computational accountability in autonomous systems, particularly in Autonomous Vehicles (AVs), where safety and efficiency often conflict. We begin by examining current approaches such as cost minimization, reward maximization, human-centered approaches, and ethical frameworks, noting their limitations addressing these challenges. Foreseeability is a central concept in tort law that limits the accountability and legal liability of an actor to a reasonable scope. Yet, current data-driven methods to determine foreseeability are rigid, ignore uncertainty, and depend on simulation data. In this work, we advocate for a new computational approach to establish foreseeability of autonomous systems based on the legal "BPL" formula. We provide open research challenges, using fully autonomous vehicles as a motivating example, and call for researchers to help autonomous systems make accountable decisions in safety-critical scenarios.

1 Introduction

Imagine an autonomous vehicle (AV) detecting a pedestrian walking in the same direction along the sidewalk with its advanced pedestrian movement prediction capabilities (Huang et al. 2022). Now, the AV must decide whether to proceed slowly behind the pedestrian, frequently stopping if it thinks the pedestrian might step into the street, or pass the pedestrian quickly to minimize delays and enhance their safety, risking a collision if the pedestrian suddenly enters the road. What should the AV do in this critical scenario?

Many autonomous systems determine actions by minimizing cost or maximizing reward (Silver et al. 2021; Wirth et al. 2017). Yet, cost minimization may not always choose an acceptable decision because of the possibility of inaction (the AV incurs no cost by remaining stationary) and the subjective nature of cost in critical situations (safety versus efficiency) if the reward system is poorly designed or not well understood which is often true in complex critical situations. Several projects, such as the Moral Machine Experiment (Awad et al. 2018) and The Molly Problem (Gal et al. 2020), use real human input to measure expectations for accountable decisions in critical situations. Furthermore, works such as (Floridi and Cowls 2022) propose principles such as beneficence, non-maleficence, autonomy, justice, and explicability to guide system decision making. Yet, the generation of these expectations and principles is often contested, incoherent, situationally dependent, inconsequential, and difficult to implement in real-world systems (Munn 2023; Huang et al. 2022). Thus, current work has left a gap between expectations and practical application in computational methods.

Although autonomous systems struggle to define acceptable decisions in critical situations, the legal system has well-established systems to ascertain reasonable decision-making for human actors. Foreseeability is a central concept in tort law that serves the essential purpose of limiting liability to a reasonable scope in negligence liability (human actors) (Zipursky 2009) and products liability (product manufacturers) (Owen 2010). Generally, a person (or entity) cannot be found negligent unless they either actually foresaw, or a reasonable person similarly situated would have foreseen, the risk of harm. Essentially, if an autonomous system is not liable for its actions, it is deemed to have made a legally acceptable choice. Thus, foreseeability can guide the decision making and evaluation of autonomous system actions in critical situations.

Currently, manufacturers use data-driven bounds defined by observed variables to assess system foreseeability for autonomous systems. These bounds establish safe behavior: if the system operates within them, events are considered foreseeable (Intelligent Transportation Systems Committee 2022). Yet, this approach faces several issues: it is rigid, ignores uncertainty, depends on simulation data, and disregards subjective factors like human behavior. Thus, the method does not effectively evaluate acceptable behaviors in real-world scenarios, as the foreseeability envelope is far narrower than the universe of computationally observable or predictable events.

One dominant way courts assess reasonable foreseeability is by using the "Hand formula" or "BPL" formula: **Burden** $<$ **Probability** $\times$ **Loss** (Posner 1972; Keating 2015; Selbst 2020). This formula considers three factors that are the probability of harm, the severity of potential harm, and the burden of precautions to prevent harm. These values are generated by a hypothetical reasonable entity in the same situation as the actor in question. If the burden of precautions

is less than the probability of harm multiplied by the potential loss caused by that harm, the situation is reasonably foreseeable and the actor could be legally liable for causing that harm. How one calculates each of those BPL factors has been the subject of much legal discourse (Gilles 2001; Schwartz 1997; Simons 2001).

In this work, we advocate for a computational approach to establish reasonable foreseeability for autonomous systems. Our proposed approach begins by establishing a library of *scenario-based reasonable standards* defining a set of acceptable behaviors reasonable for specific events to compare against real-world actions. Foreseeability is calculated using values of burden, probability, and loss based on those standards. We use a Society of Automotive Engineers level 5 (SAE International 2021) fully autonomous vehicle as a motivating example throughout the paper. However, this approach can be generally applied to any autonomous system where the system's decisions are fully made without human input.

For the rest of the paper, we will describe the proposed computational foreseeability approach and outline open research challenges, most of which require the advancement of AI technologies. We call for researchers to address these open challenges to better allow autonomous systems to make accountable decisions in safety-critical scenarios.

2 Related Work

AI accountability and ethics. Autonomous systems have several different approaches to make acceptable decisions in critical situations (Huang et al. 2022). Some approaches focus on fairness to minimize harm to various sub-groups (Feuerriegel, Dolata, and Schwabe 2020; Caton and Haas 2024), while others prioritize explainability to understand why decisions are made (Dwivedi et al. 2023). Other guiding principles include autonomy, beneficence, non-maleficence, justice, robustness, and privacy (Peters et al. 2020; Huang et al. 2022), established by government or legal organizations, including the GDPR (European Union 2018) and EU AI Act (European Parliament and Council of the European Union 2024). Yet, these methods and principles are often contested, incoherent, situationally dependent, inconsequential, and difficult to implement in real-world systems (Munn 2023; Huang et al. 2022). Thus, we suggest the use of the objective 'BPL" method to evaluate foreseeability.

Explainability for foreseeability. Explainable methods such as feature relevance, contrastive explanations, and counterfactual explanations can be used to determine foreseeability via causation (Fraser, Simcock, and Snoswell 2022). Feature relevance returns the most important features and their weights in choosing an action (Lundberg and Lee 2017; Ribeiro, Singh, and Guestrin 2016; Selvaraju et al. 2017). Contrastive explanations explain why one action is chosen over another (Stepin et al. 2021). Counterfactual explanations highlight how a change in action occurs due to a change in input (Wachter, Mittelstadt, and Russell 2017). Yet, for all these explainable methods, the determination of foreseeability is extremely subjective. Once causes are presented to the user, they must manually determine which actors are most responsible for those causes, and how to allocate liability.

Data-driven bounds for foreseeability. Currently, data-driven methods are used to determine foreseeability in autonomous systems. Manufacturers define key variables (i.e., speed, location, etc.), running simulations of specific events to generate values for those variables. If the AV operates within these limits, events are deemed foreseeable because the AVs behavior is considered safe (Intelligent Transportation Systems Committee 2022). Additional research aims to improve foreseeability by expanding or tightening these bounds using event distributions in real-world traffic data (Nakamura et al. 2022; Muslim et al. 2023). However, the use of data-driven bounds has several issues. First, they potentially overlook foreseeable events, due to their reliance solely on simulation data. Furthermore, subjective factors (i.e., human behaviors, emotions) are often ignored due to a lack of AV sensors. The generated bounds are also rigid, categorizing events as foreseeable without considering uncertainty or importance. Furthermore, the same bounds are used for all events of the same type, which may not be appropriate. Adjusting bounds based on event frequency rather than importance may overlook critical but infrequent occurrences. We suggest the use of the "BPL" method instead of data-driven bounds as it does not ignore unlikely significant features and events.

3 Open Challenges

Our suggested method (using AVs as an example system) uses reasonable AV standards to determine the burden, probability, and loss of a situation. Then, following the legally accepted formula, if burden is less than probability times loss, the situation is foreseeable; otherwise, it is unforeseeable. We can then present the ratio of burden to probability times loss to suggest the amount of foreseeability. If the AV following the reasonable standards, with its appropriate behavior, cannot handle a situation then the situation is deemed not reasonably foreseeable for all AVs, absolving the real-world AV of liability and marking its actions as acceptable in a critical situation. We present the challenges for this method below.

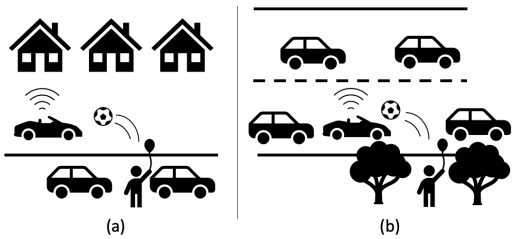

Figure 1: Running Example Scenarios. Scenario a depicts a speeding AV in quiet neighborhood where a child runs into the road after a ball. Scenario b shows the same scenario, but the AV is traveling under the speed limit on a busy highway.

Example 1 *Consider a scenario where a ball rolls into the road in a quiet residential neighborhood, unexpectedly followed by a child (Figure 1a). The AV chooses to continue*

forward, hits the child, and kills him. Was this behavior acceptable? If there was no simulation data involving the event (highly likely since the AV did not predict the child), then the data-driven bounds method indicates that the situation was unforeseeable. So, the AV's choice was acceptable. However, we know that this is untrue. All drivers should be able to anticipate a child running after a ball in a residential neighborhood.

Using the "BPL" method, we know that for a reasonable AV, the expected loss should be high even if the actual AV had no data on the event, because the magnitude of loss of a child's life is very high. Furthermore, the burden of precaution for an AV is low. It could easily slow down or stop in a quiet neighborhood. Therefore, the burden is significantly outweighed by the probability times loss, making the situation foreseeable.

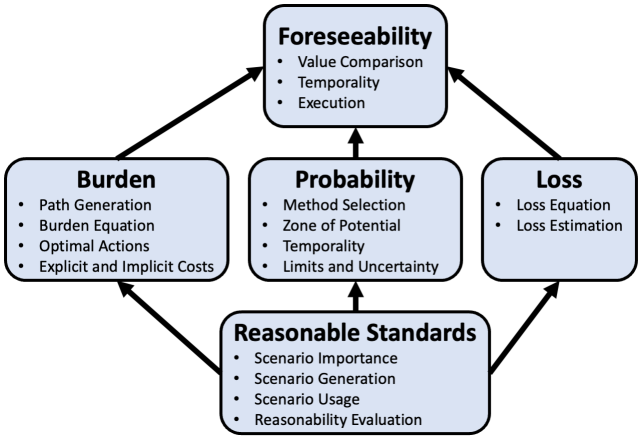

Figure 2: Map of open Challenges. These values are used to generate foreseeability from the "BPL" formula.

3.1 Reasonable Standards

Related Work. Before we can calculate foreseeability, we must generate a reasonable standard of behavior, as the values for foreseeability are based on the objectively reasonable actor rather than on real-world actions. A naive approach, regarding AVs, may be to model and predict the ordinary bounds of a human driver's perception and judgment before evaluating the system against a human's capacity. Using process-based safety and output regulation via human experts (Smith 2013), comparing individual accidents to human behavior via statistical approaches for practical accountability (Koopman and Widen 2023), and generating milestones (e.g., miles driven) through human experts are all human-based methods to evaluate AVs (Avary and Dawkins 2020).

Not only are human-centered standards vague, but comparing AVs to human drivers can be problematic. For instance, during high-speed maneuvers, AVs have greater predictive capabilities. So, they may be more culpable for a resulting accident due to their greater knowledge. Conversely, in a situation where an AV must interpret social cues, for example, a biker trying to make eye contact to cross the street, the AV may struggle more than a human driver. Therefore, comparing AVs directly to human drivers is not always appropriate. So, we must generate a standard of reasonable AV behavior, not a comparison to human driving.

Suggested Method. We recommend adapting scenario-based verification to generate individual reasonable AV standards for specific settings. First, we maintain a library of scenario-based standards defining a set of acceptable behaviors reasonable for specific events (e.g., lane change). Users can choose from existing standards in the library or generate a new one for additional scenarios. When a new set of standards is generated, an expert human (builder) builds a scenario set (data set) based on the type of event, selecting relevant input factors (e.g., number of vehicles, speed, weather, etc). Next, an expert human, or set of humans, (validator) validates that the inputs (e.g., scenarios, factors, assumptions, limits, etc.) are reasonable and complete for the given event. It is reasonable for two cars to merge at the same time during a highway lane change, for example, but not for a plane to land on the highway. The AV system then selects its actions in each scenario. If the AV's actions under the standards are deemed reasonable by the validator, then the generation terminates. If not, then the validator contacts the builder with any possible behavior issues or scenario gaps. The builder then updates the standard scenarios and AV behavior until all reasonable behaviors are captured.

Challenges. Developing a set of reasonable standards has several challenges. First, we must pinpoint critical scenarios (e.g., most common, most harmful, etc.), as it is impossible to cover an infinite number of situations. If a scenario or event type is needed, but not added in the initial library, then it can be built and added at a later time. Achieving a standard for all scenarios and event types is impossible as it would mean having a perfect AV, leaving no need for further research in this field.

Next, we must determine efficient scenario construction methods. The use of a human expert (builder) is possible but challenging as the number of scenarios and factors increases. Therefore, we recommend an automated generation system based on large language models using reported accidents, insurance claims, and legal cases supervised by a human validation expert. Still, this system must address relevance and scalability issues.

After generating scenarios, we convert the reasonable standards into a usable model to determine our values. Before addressing the question of what methods should be used (i.e. simulations, rule-based, etc.), we must understand how to use the scenarios themselves. Should all the scenarios be used or should we use only a subset of those possible? What subset should be used for testing and evaluation? Should we use only the most important time points in the scenarios? Can we generate equivalent scenarios and how many scenarios do we need in total?

Lastly, we must determine a way to evaluate the standard as reasonable. As the reasonable standards are of a reasonable AV, we cannot compare them to a human driver. However, asking human drivers to evaluate the given standard may result in a comparison to their own behavior. So, we must find a way to ensure that any behavior selected as rea-

sonable is accepted for a given AV. Furthermore, we must ensure that some reasonable behavior is selected as it could be possible no reasonable behavior can be agreed upon.

Example 2 *Let us assume a speeding AV is in an accident in a quiet residential neighborhood where a child runs after a ball into the road (Figure 1a). Reasonably, an AV should not speed in a residential neighborhood as it can cause a significant amount of danger to itself and to nearby persons and property, but it may learn to speed if told to prioritize travel time. Yet, questions remain regarding the AV's behavior, such as how closely should the AV obey the speed limit, or how quickly should the AV stop once the ball or child is seen.*

Let us assume an AV is involved in the same accident but on a busy highway (1b). A reasonable AV may slow down when a ball or child is seen on the side of the highway, but it may not be possible to travel slowly enough (or to stop) to ensure the child's safety due to the busy flow of other traffic and the need to protect its passengers from danger (e.g., tailgating, sideswiping, etc.). So, further discussion is needed here as well regarding the AV's reasonable behavior.

3.2 Burden

Related Work. The calculation of burden is closely linked to current methods for the evaluation of cost such as opportunity cost (Buchanan 1991) and multi-objective optimization (Gunantara 2018), which compare the trade-off between different choices for an agent. The survey (Wang et al. 2020) describes various methods for measuring the costs of actions in autonomous systems. Our suggested burden represents the difference between the real expected reward (or cost) and the reward (or cost) that should have been expected when acting reasonably. It is important to note even a reasonable AV may be unable to avoid all losses. In such cases, a low burden may suggest accepting minor losses, while a high burden may indicate an unforeseeable outcome.

Suggested Method. Burden refers to the cost incurred by the AV to prevent a situation, involving the reasonable AV standards, the real-world AV model, and the available situation data. We suggest the following method for calculating the burden for the given situation. From the determined situation start, we isolate the agent's optimal chain of actions (or path) from the real-world AV's policy and reward structure leading to the situation, generating a maximum cumulative expected reward. This reward represents the reward the AV expected to receive. Next, we isolate the optimal chain of actions (or path) from the reasonable AV standards for the given situation, generating a second maximum cumulative expected reward. This represents the path that the real-world AV should have taken and the reward it should have received regardless of its training. Finally, subtract the real-world AV's maximum reward from the reasonable AV's maximum reward to compute the burden of the untaken precaution.

Challenges. Calculating burden has several challenges. First, generating expected cumulative rewards from the given models for the needed paths is difficult due to the infinite number of factor combinations, the possibility of uncertainty (e.g., other road users, weather, etc.), and the black-box nature of the models. Furthermore, the evaluation of rewards involves a combination of many different factors (e.g., time, safety, comfort, efficiency, etc.). So, the most appropriate factors need to be chosen for the given AV and situation.

Additionally, this expected reward lies in generating the optimal actions for the real-world AV which may be subjective and unclear before (or even after) the situation occurs. To generate the optimal actions, explainability methods such as contrastive or counterfactual explanations (Stepin et al. 2021; Wachter, Mittelstadt, and Russell 2017) may be used, exploring why one behavior occurs over another or how changes in output occur based on a behavior change.

Finally, the burden equation remains in question. Our suggested method is based on opportunity cost (i.e., the loss of reward from one action when another action is chosen) (Buchanan 1991), comparing the best two paths for the available models. However, it may be more appropriate to use the expected reward from the average of all possible paths, the expected reward from the worst possible paths, or another method entirely. Furthermore, we need to consider not just the explicit costs to the AV which it can measure using its sensors (e.g., time to reach the AV's destination), but also implicit or outside costs (e.g., depreciation of the vehicle) which can be difficult to capture. Overall, further exploration is needed.

Example 3 *Looking at Table 1, for our neighborhood scenario, we find the generated burden is quite small. The real-world AV determines the original cost of the scenario to be 9,000 due to travel time and safety requirements. However, the reasonable AV suggests slowing down and stopping abruptly when the ball is seen, prioritizing the child's safety over completion time, passenger comfort, and efficiency. This increases the cost to 37,000. So, the overall burden is 28,000. For our highway scenario, the burden is significantly larger. The real-world AV determines the cost of the scenario to be 36,000. Yet, stopping abruptly on the highway could cause a slight delay in time, comfort, and efficiency, but it will also cause the following vehicle to rear-end the AV, potentially causing injury or death of its passengers. The potential of serious injury or death increases the cost significantly to 1,914,000. So, the overall burden is 1,878,000.*

3.3 Probability

Related Work. Methods to generate probability are well defined. Scene-graphs (Malawade et al. 2022), LSTMs (Zhang and Abdel-Aty 2022), decision trees/random forests (Nahata et al. 2021), MDPs (Althoff, Stursberg, and Buss 2009), and model predictive control (Wang et al. 2019) have all been used to predict the probability of AV accidents. Furthermore, surveys like (Huang et al. 2022) provide further methods for probability generation such as physics models, Monte Carlo methods, machine learning, deep learning, and reinforcement learning.

Suggested Method. Probability determines the likelihood of the situation using the reasonable AV standards and the situation data from the real-world AV. To generate proba-

	Neighborhood	Highway
Burden	|-9,000 - (-37,000)| = **28,000**	|-36,000 - (-1,914,000)| = **1,878,000**
Probability	1 second - Fatal Injury (100%) 2 seconds - Fatal Injury (80%), Serious Injury (20%) **3 seconds - Fatal Injury (60%), Serious Injury (25%), No Accident (15%)** 4 seconds - No Accident (100%)	1 second - Fatal Injury (100%) **2 seconds - Fatal Injury (45%), Serious Injury (15%), No Accident (40%)** 3 seconds - Fatal Injury (20%), No Accident (80%) 4 seconds - No Accident (100%)
Loss	(0.15 * 7,100) + (0.25 * 162,000) + (0.6 * 1,869,000) = **1,162,965**	(0.4 * 7,100) + (0.15 * 162,000) + (0.45 * 1,869,000) = **868,190**
Foreseeability	28,000 < 1,162,965 → True → **Foreseeable**	1,878,000 < 868,190 → False → **Unforeseeable**

Table 1: Calculations of burden, probability, loss, and foreseeability for running example scenarios shown in Figure 1.

bility, we must first determine the situation start state and situation length from the real-world situation data. Then, we determine all reachable states within the situation length and calculate the probability of reaching a failure state within the generated paths for the reasonable AV.

Challenges. However, generating probability for foreseeability presents several specific challenges. First, we must determine the best way to generate the probability output. Should we use historical data, simulations, and frequency counting from the exact given situation, or a more complex neural network such as an LSTM? We must compare the current state-of-the-art methods in collaboration with AV and legal experts.

Second, we must determine the actual start state for prediction from the larger collection of AV data. Event data recorders (black boxes) may record different variables at different rates depending on vehicle sensors (Chidester, Hinch, and Roston 2001). So, we must determine when to generate our probability value (e.g., when the AV is turned on, average black box recording, when the AV still has time to act, etc.). Determining this "zone of potential" is essential, a gray area between extremes where the AV can potentially prevent the situation. The zone must not be so early that the AV cannot accurately predict the possibility of a situation, but not so late that the AV cannot do anything to prevent the situation. This "zone of potential" must be set for each AV and situation.

Conversely, we can show how probability and foreseeability change as the AV makes decisions over time. Yet, we must decide when and how often foreseeability is calculated. Furthermore, we must determine how to present this information to the user efficiently and understandably.

Finally, we must determine any limits or uncertainty in the generated probability. We may want to limit actions, paths, or factors in our probability generation if they are too distant from the original situation, or irrelevant to its outcomes. As the situation may change over time, we must constantly manage the influence of these limits. Furthermore, we must account for any relevant uncertainty. Specifically, changes in environmental factors that are possible but uncontrollable (e.g., other aggressive drivers, weather, etc.). By developing a set of diverse, but relevant actions, paths, and factors, the method can come to a more accurate probability of situation outcomes.

Example 4 *Returning to our example, running our probability algorithm yields results for seconds before the accident shown in Table 1. In the neighborhood scenario, we choose to generate probability three seconds before the accident as four seconds prior the AV cannot predict the accident, and two seconds prior it cannot avoid it. For our highway example, we select to generate our probability two seconds before the accident as it represents the last moment the accident could be avoided once its possibility is detected. Four seconds prior the AV does not predict the accident as the ball is not seen. Three seconds prior the AV sees the ball but does not predict an accident as the ball could have come from the car in front of the AV. A child is unlikely. One second before, the accident is unavoidable.*

3.4 Loss

Related Work. Loss can be determined from several different sources, such as evaluations of tort law (Sugarman 2015; Avraham 2006), state and federal payment caps (Legal Information Institute 1999), legal cases (CBS News 2021), insurance payouts (Insurance Information Institute 2022), and cost analysis from data surveys (National Safety Council 2022). However, no one source is better than another as it is very difficult to place a value on human life, damage, pain, and suffering. Domain experts must choose the best loss estimate based on individual autonomous systems and situations.

Suggested Method. Loss represents the amount of harm or damage that a situation causes. We can generate loss using real-world situation data and data regarding insurance or legal case payouts. We suggest using insurance or legal case data to create a table of costs (e.g., financial, physical, emotional, etc.) per outcome. For each possible situation outcome, we determine a matching table entry and pull the cost value. The combined loss and probability are then determined by multiplying the cost by the probability of the outcome. Adding these products together results in the total combined probability and loss.

Challenges. Domain experts must answer questions such as whether all situations of a similar type should be evaluated for the same loss, whether loss should be weighted based on situational factors, and whether loss should contain only easily measured objective loss for the AV or should include subjective loss as well.

Relatedly, how should loss be estimated for nontrivial situations? If there is an exact match in the generated table (common situation type), then it can be retrieved. However, for more unique and complex situations, an estimation of more general loss from known insurance or legal data is needed.

Example 5 *For our running example, we will use the National Safety Council (NSC) average economic cost of motor vehicle crashes by injury severity (National Safety Council 2022). According to the NSC, fatal injuries cost $1,869,000, serious injuries cost $162,000, minor injuries cost $42,000, possible injuries cost $26,000, and no injury costs $7,100. Table 1 shows the calculation of loss for both examples.*

3.5 Foreseeability

Suggested Method. As a matter of law, reasonable foreseeability is determined based on the "BPL" formula. Here, we compare the total combined probability and loss to the burden. If the burden is lower than the combined probability and loss, then the situation is reasonably foreseeable; otherwise, it is not. We can also compare the ratio of the combined probability and loss to the burden to get an understanding of how foreseeable or unforeseeable the situation is.

Challenges. It still can be challenging to determine foreseeability even with the values of burden, probability, and loss already determined. First, we must make sure that the values are normalized so that they can be accurately compared. For example, linking the cost of loss with the reward structure in the generation of burden ensures a 1-to-1 comparison. Providing users with information on how these values were calculated is crucial for an effective understanding of resulting foreseeability.

We must consider the temporality of the given situation. If generating foreseeability over time, we need to decide how often to calculate it and how to present changes understandably. However, when regarding foreseeability as a whole, not just probability, we must consider how our choice of "zone of potential" will influence the values of burden, probability, and loss. For example, four seconds before the situation, the AV may predict a lower probability of the situation compared to three seconds prior, but there will likely be less burden on the AV to prevent the situation and less potential loss if better actions are chosen earlier on. So, when the values are calculated, they must be weighted appropriately to generate an accurate estimate of foreseeability.

Finally, we need to know when to calculate foreseeability. Should it be event-driven (when a situation occurs) or periodical (consistent over time)? This choice may depend on the automation level of the AV, user preference, and situation type. However, care should be taken not to overwhelm the user, the AV, or the system.

Example 6 *Comparing our generated values found in Table 1, we find our neighborhood scenario reasonably foreseeable. This is expected since the AV can avoid the accident by just being more cautious and slowing down. However, the highway scenario is not reasonably foreseeable, due to the lower probability of the accident (a child running onto the highway is less likely than a neighborhood street) and the high burden (due to traffic, slowing down quickly will cause further accidents and the passenger's death). Thus, the AV in the neighborhood scenario makes an unacceptable decision, whereas the AV in the highway scenario makes an acceptable one.*

4 Advancing Towards Real-World Solutions

We now present barriers to advancing research on foreseeability implementation and utilization in real-world systems.

Data. There is a lack of crucial data to determine foreseeability. While some data exists for AVs and other autonomous systems, such as (Chen et al. 2024; Agarwal et al. 2020), which can be used to predict situational outcomes, this data lacks features and labeling needed for foreseeability. However, how should these annotations occur? Who should provide them? How do we resolve any labeling conflicts due to disparate opinions on foreseeability? Finally, the large number of possible complex critical scenarios makes compiling a comprehensive data set difficult. Yet, it is possible to utilize datasets of manual vehicle accidents (e.g., (Moosavi et al. 2019a,b)) to generate additional examples for autonomous vehicles by simulating AV behavior based on the scenarios in the dataset. So, we must collaborate with autonomous system manufacturers and legal experts to collect quality data.

Methodology. Although there has been significant growth in methods for developing autonomous systems, the context of foreseeability has several distinct challenges. Due to situational complexity, selecting a relevant subset of factors to determine foreseeability may be difficult, especially when subjective or situational factors are present. Furthermore, the complexity could lead to a large number of parameters that must be selected and optimized. Specific domain knowledge may also be needed to address these problems and identify issues or biases with produced methods. Finally, what kinds of current methods should be used? Models such as LSTMs (Zhang and Abdel-Aty 2022) and LLMs (Cui et al. 2024) could both be used to analyze case descriptions or environmental pictures (i.e., traffic cameras) to produce "BPL" values. However, each method must be analyzed and tested for possible effectiveness, complications, and drawbacks.

Evaluation. Finally, we must produce a complete and effective evaluation framework. Due to the lack of data, any computational evaluation of generated methods suffers from an unknown ground truth. Furthermore, when data is produced, limited examples would be available for evaluation due to the small number of reported accidents involving autonomous systems (Wansley 2021; Waymo 2024) and current lack of standardization and implementation of AI regulation (Munn 2023). Evaluation with real-world users faces similar problems. Which test groups should be used (e.g., legal professionals, AV experts), if each group has its own opinion of acceptable behavior? Furthermore, motivating and ensuring users evaluate systems without bias towards human actions is difficult.

5 Conclusion

Overall, we call for the establishment of a computational foreseeability framework to guide the evaluation of autonomous system decision making in critical situations. We call for researchers to answer these open challenges to improve autonomous systems, as the extension of current work can build a basis for safe, responsible use of these technologies.

Acknowledgments

This work was supported in part by U.S. National Science Foundation under grants CCF-2131511 and CCF-2131531. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the grant sponsors.

References

Agarwal, S.; Vora, A.; Pandey, G.; Williams, W.; Kourous, H.; and McBride, J. 2020. Ford multi-AV seasonal dataset. *The International Journal of Robotics Research*, 39(12): 1367–1376.

Althoff, M.; Stursberg, O.; and Buss, M. 2009. Model-based probabilistic collision detection in autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 10(2): 299–310.

Avary, M.; and Dawkins, T. 2020. Safe Drive Initiative: creating safe autonomous vehicle policy. In *World Economic Forum: Geneva, Switzerland*.

Avraham, R. 2006. Putting a price on pain-and-suffering damages: A critique of the current approaches and a preliminary proposal for change. *Northwestern University Law Review*, 100: 87.

Awad, E.; Dsouza, S.; Kim, R.; Schulz, J.; Henrich, J.; Shariff, A.; Bonnefon, J.-F.; and Rahwan, I. 2018. The Moral Machine experiment. *Nature*, 563(7729): 59–64.

Buchanan, J. M. 1991. Opportunity Cost. In Eatwell, J.; Milgate, M.; and Newman, P., eds., *The World of Economics*, 520–525. London: Palgrave Macmillan.

Caton, S.; and Haas, C. 2024. Fairness in machine learning: A survey. *ACM Computing Surveys*, 56(7): 1–38.

CBS News. 2021. Tesla settles lawsuit with family of man killed in Autopilot crash. https://www.cbsnews.com/sanfrancisco/news/tesla-autopilot-fatal-crash-lawsuit-settlement-walter-huang-apple-engineer-mountain-view/.

Chen, K.; Ge, R.; Qiu, H.; Ai-Rfou, R.; Qi, C. R.; Zhou, X.; Yang, Z.; Ettinger, S.; Sun, P.; Leng, Z.; Mustafa, M.; Bogun, I.; Wang, W.; Tan, M.; and Anguelov, D. 2024. WOMD-LiDAR: Raw Sensor Dataset Benchmark for Motion Forecasting. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.

Chidester, A.; Hinch, J.; and Roston, T. A. 2001. Real world experience with event data recorders. In *Proceedings of the Seventeenth International Technical Conference on the Enhanced Safety of Vehicles, Amsterdam, Netherlands (June 2001)*.

Cui, C.; Ma, Y.; Cao, X.; Ye, W.; Zhou, Y.; Liang, K.; Chen, J.; Lu, J.; Yang, Z.; Liao, K.-D.; et al. 2024. A survey on multimodal large language models for autonomous driving. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 958–979.

Dwivedi, R.; Dave, D.; Naik, H.; Singhal, S.; Omer, R.; Patel, P.; Qian, B.; Wen, Z.; Shah, T.; Morgan, G.; et al. 2023. Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, 55(9): 1–33.

European Parliament and Council of the European Union. 2024. EU Artificial Intelligence Act. Official Journal of the European Union.

European Union. 2018. General Data Protection Regulation (GDPR). https://gdpr-info.eu/.

Feuerriegel, S.; Dolata, M.; and Schwabe, G. 2020. Fair AI: Challenges and opportunities. *Business & information systems engineering*, 62: 379–384.

Floridi, L.; and Cowls, J. 2022. A unified framework of five principles for AI in society. *Machine learning and the city: Applications in architecture and urban design*, 535–545.

Fraser, H.; Simcock, R.; and Snoswell, A. J. 2022. Ai opacity and explainability in tort litigation. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 185–196.

Gal, D.; Iorio, L.; Krugel, S.; Dixon, L.; Uhl, M.; and Dawkins, T. 2020. AI Safety & Ethics for Self-Driving: Introducing the Molly Problem. https://aiforgood.itu.int/event/ai-safety-ethics-for-self-driving-introducing-the-molly-problem/.

Gilles, S. G. 2001. On Determining Negligence. *Vanderbilt Law Review*, 54: 813–861.

Gunantara, N. 2018. A review of multi-objective optimization: Methods and its applications. *Cogent Engineering*, 5(1): 1502242.

Huang, Y.; Du, J.; Yang, Z.; Zhou, Z.; Zhang, L.; and Chen, H. 2022. A survey on trajectory-prediction methods for autonomous driving. *IEEE Transactions on Intelligent Vehicles*, 7(3): 652–674.

Insurance Information Institute. 2022. Facts + Statistics: Auto insurance. https://www.iii.org/fact-statistic/facts-statistics-auto-insurance#Passenger

Intelligent Transportation Systems Committee. 2022. IEEE Standard for Assumptions in Safety-Related Models for Automated Driving Systems. *IEEE Std 2846-2022*, 1–59.

Keating, G. C. 2015. Must the Hand Formula Not Be Named? *University of Pennsylvania Law Review Online*, 163: 367–375.

Koopman, P.; and Widen, W. H. 2023. A Reasonable Driver Standard for Automated Vehicle Safety. In *International Conference on Computer Safety, Reliability, and Security*, 355–361. Springer.

Legal Information Institute. 1999. 15 U.S. Code § 6604. https://www.law.cornell.edu/uscode/text/15/6604.

Lundberg, S. M.; and Lee, S.-I. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

Malawade, A. V.; Yu, S.-Y.; Hsu, B.; Muthirayan, D.; Khargonekar, P. P.; and Al Faruque, M. A. 2022. Spatiotemporal scene-graph embedding for autonomous vehicle collision prediction. *IEEE Internet of Things Journal*, 9(12): 9379–9388.

Moosavi, S.; Samavatian, M. H.; Parthasarathy, S.; and Ramnath, R. 2019a. A countrywide traffic accident dataset. *arXiv preprint arXiv:1906.05409*.

Moosavi, S.; Samavatian, M. H.; Parthasarathy, S.; Teodorescu, R.; and Ramnath, R. 2019b. Accident risk prediction based on heterogeneous sparse data: New dataset and insights. In *Proceedings of the 27th ACM SIGSPATIAL international conference on advances in geographic information systems*, 33–42.

Munn, L. 2023. The uselessness of AI ethics. *AI and Ethics*, 3(3): 869–877.

Muslim, H.; Endo, S.; Imanaga, H.; Kitajima, S.; Uchida, N.; Kitahara, E.; Ozawa, K.; Sato, H.; and Nakamura, H. 2023. Cut-out scenario generation with reasonability foreseeable parameter range from real highway dataset for autonomous vehicle assessment. *IEEE Access*.

Nahata, R.; Omeiza, D.; Howard, R.; and Kunze, L. 2021. Assessing and explaining collision risk in dynamic environments for autonomous driving safety. In *2021 IEEE international intelligent transportation systems conference (ITSC)*, 223–230. IEEE.

Nakamura, H.; Muslim, H.; Kato, R.; Préfontaine-Watanabe, S.; Nakamura, H.; Kaneko, H.; Imanaga, H.; Antona-Makoshi, J.; Kitajima, S.; Uchida, N.; et al. 2022. Defining reasonably foreseeable parameter ranges using real-world traffic data for scenario-based safety assessment of automated vehicles. *IEEE Access*, 10: 37743–37760.

National Safety Council. 2022. Injury Facts. https://injuryfacts.nsc.org/all-injuries/overview/.

Owen, D. G. 2010. Bending Nature, Bending Law. *Florida Law Review*, 62: 569–615.

Peters, D.; Vold, K.; Robinson, D.; and Calvo, R. A. 2020. Responsible AI—two frameworks for ethical design practice. *IEEE Transactions on Technology and Society*, 1(1): 34–47.

Posner, R. A. 1972. A Theory of Negligence. *Journal of Legal Studies*, 1: 29–96.

Ribeiro, M. T.; Singh, S.; and Guestrin, C. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 1135–1144.

SAE International. 2021. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. https://www.sae.org/standards/content/j3016_202104/.

Schwartz, G. T. 1997. Mixed Theories of Tort Law: Affirming Both Deterrence and Corrective Justice. *Texas Law Review*, 75: 1801–1834.

Selbst, A. D. 2020. Negligence and AI's Human Users. *Boston University Law Review*, 100: 1315–1376.

Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; and Batra, D. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 618–626.

Silver, D.; Singh, S.; Precup, D.; and Sutton, R. S. 2021. Reward is enough. *Artificial Intelligence*, 299: 103535.

Simons, K. W. 2001. The Hand Formula in the Draft Restatement (Third) of Torts: Encompassing Fairness as Well as Efficiency Values. *Vanderbilt Law Review*, 54: 901–939.

Smith, B. W. 2013. The Reasonable Self-Driving Car. Center for Internet and Society at Stanford Law School.

Stepin, I.; Alonso, J. M.; Catala, A.; and Pereira-Fariña, M. 2021. A survey of contrastive and counterfactual explanation generation methods for explainable artificial intelligence. *IEEE Access*, 9: 11974–12001.

Sugarman, S. D. 2015. Tort damages for non-economic losses: Personal injury. In *Comparative Tort Law*, 323–356. Edward Elgar Publishing.

Wachter, S.; Mittelstadt, B.; and Russell, C. 2017. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31: 841.

Wang, H.; Huang, Y.; Khajepour, A.; Zhang, Y.; Rasekhipour, Y.; and Cao, D. 2019. Crash mitigation in motion planning for autonomous vehicles. *IEEE transactions on intelligent transportation systems*, 20(9): 3313–3323.

Wang, Q.; Ma, Y.; Zhao, K.; and Tian, Y. 2020. A comprehensive survey of loss functions in machine learning. *Annals of Data Science*, 1–26.

Wansley, M. 2021. The End of Accidents. *U.C. Davis Law Review*, 55: 269–345.

Waymo. 2024. Safety Impact. https://waymo.com/safety/impact/.

Wirth, C.; Akrour, R.; Neumann, G.; and Fürnkranz, J. 2017. A survey of preference-based reinforcement learning methods. *Journal of Machine Learning Research*, 18(136): 1–46.

Zhang, S.; and Abdel-Aty, M. 2022. Real-time crash potential prediction on freeways using connected vehicle data. *Analytic methods in accident research*, 36: 100239.

Zipursky, B. C. 2009. Foreseeability in Breach, Duty, and Proximate Cause. *Wake Forest Law Review*, 44: 1247–1275.