Maximizing Safety: Smart Autonomous Path Planning to Reduce Harm from Crashes

AP Research

Word Count: 2594

Author Note

This experimental study and research report were completed in partial fulfillment of the requirements for AP Research.

Abstract

This project seeks to minimize damages caused in car accidents involving one or more autonomous vehicles due to the high economic and medical costs of traffic collisions. When accounting for societal harm, car crashes cost the US $1.4 trillion in 2019 alone. To accomplish this, an autonomous vehicle control algorithm based on an ensemble model that employs weighted majority voting was developed, which uses variable weights for each base learner depending on the likelihood of a crash occurring. Feedback from sensors and trajectory prediction was used to determine the likelihood of a crash. Damage was defined as the sum of medical costs to the driver and repair costs for the vehicle and the environment. I compared the damages from my model and a high-performance open-source model using the simulator CARLA, where each model was tested five times per map on eight different maps. A Bayesian hierarchical model constructed with the data from testing each model estimated a 13% reduction in damage, with a 91% posterior probability that the reduction was attributable to the created algorithm. These results demonstrate the use case of ensemble models in autonomous driving algorithms and their ability to reduce damages caused in car crashes. Future research could evaluate alternative weighting metrics or compare parallel and sequential ensemble architectures.

Introduction

The development of Level 4 autonomous vehicles is driven by the goal to improve road safety by removing human error. To achieve this, research has been done to support the advancement of trajectory prediction and path planning technologies (Cho et al., 2020). Researchers have also optimized algorithms to detect environmental risks to calculate collision-free routes (Zhang et al., 2023). In ideal conditions, these systems are capable of navigating complex environments with precision.

However, a foundational limitation remains in the logic of current autonomous navigation. The assumption that a safe path always exists is made, but in the chaotic reality of traffic, scenarios inevitably arise where a collision is mathematically unavoidable due to sudden environmental changes or the erratic behavior of other vehicles. Current path-planning paradigms are designed almost exclusively for avoidance; when avoidance becomes impossible, these models often lack the decision-making framework to distinguish between a minor fender-bender and a catastrophic impact. Consequently, an AV facing an inevitable crash may default to suboptimal behaviors that fail to mitigate the severity of the outcome.

This research addresses the disconnect between the unachievable, binary goal of collision avoidance and the feasible, nuanced reality of damage minimization. While theoretical discussions regarding the ethics of inevitable accidents exist, there is a significant lack of practical, applied algorithms designed to handle these edge cases. This study seeks to bridge that gap by developing a trajectory planning model specifically for inevitable collision scenarios. By utilizing the CARLA simulator to compare traditional avoidance models against a novel algorithm optimized to minimize monetary harm, defined as a composite of injury severity and property damage, this research evaluates whether a system designed to fail safely can statistically reduce the aggregate cost of unavoidable accidents.

Literature Review

Relevant Terminology

Autonomous path planning is the process whereby a self-driving vehicle determines a route based on its environment. This is done through the system prioritizing certain factors, generally being safety and efficiency, alongside information relating to the environment. With regards to autonomous vehicles, the environment refers not only to nature, roads, and other structures but also includes people, vehicles, and other objects nearby. Trajectory prediction is used to forecast how objects around the vehicle will act and is of imperative importance for effective autonomous path planning. As these systems grow more capable, their ability to make change increases. The National Highway Traffic Safety Administration (NHTSA, 2015) found that the critical reason for a crash was assigned to the driver in 94% of the crashes it investigated, showcasing the extreme importance of effective autonomous driving systems. With effective systems, the damages and harm caused by traffic accidents can drop significantly, saving many lives and reducing much monetary damage in the process. This study focuses on multi-agent driving environments involving Level 4 autonomous vehicles, where trajectory prediction fails to produce a collision-free path. Monetary damage is defined through predicted injury severity metrics and predicted infrastructure and vehicle damage metrics.

Existing Research

Autonomous driving systems rely on the ability to create paths that take into account immediate conditions and also predicted changes in the environment. Previous works have used methods such as probabilistic reasoning, deep learning, and reinforcement learning to create and refine the accuracy of the models used to generate these paths. Cho et al. (2020) established foundational principles such as the importance of deterministic motion modeling and simultaneous localization and mapping (SLAM) for trajectory creation. This method allows the vehicle to identify its surroundings while learning where it is in relation to these objects.

However, these methods proved to be unreliable when tested in chaotic or unpredictable environments because of the increase in computational expenses when using SLAM in these environments. Zhang et al. (2023) helped to address this limitation by introducing probabilistic semantic mapping, which helped resolve many of the problems faced by the approach taken by Cho et al. Probabilistic semantic mapping identifies objects by breaking them down into pieces and analyzing each one, where the granularity can be adjusted depending on the scenario, allowing for a balanced computational expense. These methods attempt to create paths within restrictions, such as physical limitations of the vehicle; however, they fail to consider passenger safety as a measure to determine the best path in circumstances where no safe path exists. They simply try to create the best path based on the risk of a collision, without considering the severity of such an impact.

Despite the growth in actual trajectory creation, many of these methods rely on outdated methods for environmental prediction, which limit their real-time accuracy. To remedy this issue, Chen et al. (2025) developed a graph-temporal convolutional network (GTC-DAN) that captured interactions among vehicles, improving predictive accuracy by thirty-six percent. GTC-DAN allowed for the incorporation of a multi-vehicle system where decisions were dependent on all external vehicles rather than treating them separately as distinct entities. This improves trajectory prediction by incorporating changes in trajectories of individual vehicles caused by reactions to other vehicles.Building off the work of Chen and his team, Zhou et al. (2025) strengthened automated scenario testing (AST) by employing a more varied dataset, improving the adaptability of environmental prediction systems. By increasing the effectiveness of automatic training systems, trajectory prediction models can be trained on a wider variety of potential scenarios to increase the likelihood that the model will not need to extrapolate when faced with a real-world scenario. The increase in training is made feasible by the reduction in time and cost of training. This is caused by the aforementioned improvements in AST. These methods both showcase the shift from rule-based predictions to data-driven, nuanced prediction methods while highlighting ways to improve these new prediction methods.

Gap in Research

However, all of the noted approaches make the same oversight; they assume that it is possible to avoid a crash with the given constraints and do not take into account when such a path is impossible to find. Although such systems have been proposed among ethical discussions, such as Xiao (2022), who noted that safety should be a main factor in decision-making processes through more nuanced path selection algorithms, there have been no practical applications or even outlines on how such a combined system could function, with analysis on its strengths and weaknesses. This leads to a gap between theoretical propositions and applied research.

If this gap persists, taxpayers and those involved in collisions stand to lose. Taxpayer dollars are wasted on repairing infrastructure, whose damage could be minimized by addressing the aforementioned gap. Additionally, those involved in crashes will face less severe injuries if this gap is addressed, reducing financial burdens and the risk of death.

This is a practice gap, because there are proposals, but no practical implementation of a system that takes into account when crashes are inevitable in trajectory decision making. My research, which attempts to create an algorithm that takes into account when crashes are inevitable to create new paths that prioritize passenger safety, would remedy this gap by providing a study that can connect current methods of trajectory generation, dependent on new and improved methods of environmental prediction, to theorize ideas to improve passenger safety and would be able to analyze the effectiveness of such a model, determining both strengths and weaknesses of the new system, tested with simulation and trained to use ethical decision making, compared to traditional ones.

Method

Study Methodology

To create a solution to fix this gap, the study identifies a causal relationship to prove that the decrease in damage during the crash was caused by the change in model used. To identify a causal relationship, the study uses an experiment. The independent variable in this study will be the model used. This study will involve the development of multiple different autonomous driving models that use varying parameter levels and take different approaches to reducing harm from crashes. Each model will then be run through CARLA, which is a simulator that can run autonomous vehicle programs, and test their performance with monetary damage as the dependent variable. The dependent variable will be tracked through the information CARLA provides, and each model will be tested in the same set of varying conditions to evaluate each model equally while having enough information to support a claim about statistically significant differences.

A quantitative method especially fits this research design because this study uses a simulation, which allows the retrieval of quantitative information about all conditions related to the performance of each model. This allows for the study to utilize a wide variety of data. This study aims to mitigate harm, which is easily converted into a numerical value by receiving damage price from the simulation and combining that with the average medical price per injury, with injuries also found from the simulation. As for the average costs, such information is published by OSHA, the Occupational Safety and Health Administration.

The study will use a simulation named CARLA. CARLA is a simulation that allows the user to create specific scenarios and test out the results of different models to control the self-driving vehicle. CARLA can be used to simulate numerous scenarios for each of the created models, recording information about the impact of the crash to allow for the calculation of monetary harm for each simulation. This monetary damage is the needed information, and it can be calculated using data from CARLA. This study will use the aforementioned strategy of information published by OSHA to convert the information given by CARLA into a number that represents the monetary harm that occurred in the simulation and link that to the model with which the simulation was run.

Study Procedure

I will follow the following steps to conduct my research and analyze the resulting data. Firstly, different models will be created, basing them on existing precedents, accepted methods, and industry norms. Then, a variety of different simulation cases in CARLA will be created to ensure enough data is gathered. Each model will be tested with each scenario, recording information that will be used to calculate monetary harm. This information includes location and angle of impact, velocity vectors of each vehicle at impact, damage to each vehicle, and acceleration vectors of each vehicle at impact, among other information. Information from each simulation will be consolidated into a single monetary harm value by using OSHA to calculate harm to people while using the physical damages given by the simulation.

Once the data has been collected, it will be analyzed using a Bayesian Hierarchical Model. The data will first be sorted based on the model, and then by simulation, with an associated harm value. A prior distribution for σ^2 which models ε will then be created. A prior distribution for σ^2 can be created using an inverse Gaussian model, allowing us to use variances α_σ and β_σ as parameters of the inverse Gaussian distribution to define σ^2. Harm will be decomposed into its simulation, model, and noise effect components. Harm for a model m and simulation s, with a recorded harm of h, will follow the following equation. $h_{ms}=\alpha_s+\beta_m+\varepsilon_{ms}$ where alpha is the variance caused by the simulation itself, beta is the variance caused by the model, and epsilon is the variance caused by errors in the computation of the harm. Epsilon is assumed to be normally distributed and centered at 0, with a variance of σ^2. The calculated values of alpha and beta will then be used to identify the variances, assuming that they are normally distributed, using the equations $\alpha_i \sim N(0, \tau_\alpha^2)$ and $\beta_j \sim N(0, \tau_\beta^2)$. These equations create a normal distribution to model the values found by step 2. τ_α and τ_β represent the variance in the simulations and models, respectively. Once this is complete, Gibbs Sampling will be used to calculate the posterior distribution. The posterior distribution can be defined as $P(\alpha, \beta, \sigma^2 \mid h)$, which calculates adjusted probabilities for alpha, beta, and sigma squared through Bayesian inference. The software MATLAB will be used to conduct the Gibbs Sampling because it is too time-consuming and computationally difficult to do manually. Once the posterior distribution is created, the posterior mean will be evaluated, and a 99% confidence interval will be created to identify the minimization of beta. This averages all potential values of beta for each model and creates an interval where these values work with 99% certainty, allowing us to complete our posterior analysis. The exact math used to evaluate the data can be found in Appendix A1, while the code used can be found in Appendix A2. Once the analysis is completed, a visualization of the post-analysis data will be created. Because a Bayesian Hierarchical Model is used, each model can be ranked by its probability of being the best, which can be shown in diagrams like a bar chart.

Results

Following the completion of all trials, monetary harm values were calculated for each trial, sorted by map, trial number, and the model that was used. A data table of all collected samples can be found in Appendix B1. It was found that the reduction in damage caused by the ensemble model was proportional to the average damage caused. A graph of the percent difference in damage between the two models per map can be found in Appendix B2.

Discussion

Findings

After completion of the aforementioned analysis, it was found that the created ensemble model outperformed the single model algorithm by approximately 13%, with a 91% posterior probability. This supports the fact that the use of an ensemble model reduces the damage caused by autonomous vehicles.

Implications

This research is significant because it has the potential to influence future work on autonomous vehicles. It has been shown that the use of ensemble models can outperform single models. This can mean that more researchers can begin to utilize ensemble models rather than single models to conduct their research. This can also shift how large companies develop their autonomous driving models because there is a clear and significant improvement when ensemble models are used, which could be used to improve their self-driving performance. This research can also spark ethical debates in the larger academic conversation, as only a few foundational sources exist for the ethicality of damage-based artificial intelligence.

Limitations

There are a few limitations to this research. Firstly is the fact that testing occurred within a simulation. Even though CARLA provides high-quality data and has been used by numerous other researchers, it cannot truly replicate real-world conditions and randomness. The second limitation is that the calculated monetary harm metric relies on estimating costs for injuries with published OSHA information, which fails to consider special characteristics that may lead to different damage values. Lastly, there is the way the statistical model accounts for external variables. According to the central limit theorem, the sample size is large enough to assume the data varies normally. However, if real-world data is not varied normally, this model will have slightly biased results because its probabilities do not accurately reflect the real world.

Conclusion and Future Research

This research investigated how the performance of an ensemble model with variances based on crash likelihood would compare to the performance of a standard, single model algorithm with regard to the total damage caused to both people and property. Using simulations to gather data, analysis with a Bayesian Hierarchical Model revealed a 13% decrease in total damage caused when the ensemble model was used, with a 91% chance that the improvement was attributable to the model itself rather than external variables. These findings provide evidence for the fact that decision-making that takes into account the chance of a crash occurring leads to a significant decrease in the damage caused through crashes. Further work in this field could integrate custom passenger setups and vehicle details to better estimate the damage caused to people in a scenario and have that damage be for the specific passenger and vehicle scenario rather than generalized.

References

Chen, H., Wu, X., Zhang, R., Guo, W., Chen, Y., Xu, J., Zhang, W., & Yu, W. (2025). GTC-DAN: A graph-temporal convolutional model with dynamic adjacency for vehicle trajectory prediction. *Bulletin of the Polish Academy of Sciences: Technical Sciences, 73*(2), Article e152610. https://doi.org/10.24425/bpasts.2024.152610

Cho, S., Kim, C., Park, J., Sunwoo, M., & Jo, K. (2020). Semantic point cloud mapping of LiDAR based on probabilistic uncertainty modeling for autonomous driving. *Sensors, 20*(20), Article 5900. https://doi.org/10.3390/s20205900

National Highway Traffic Safety Administration. (2015). *Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey* (Report No. DOT HS 812 115). U.S. Department of Transportation. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115

Xiao, Y. (2022). Application of machine learning in ethical design of autonomous driving crash algorithms. *Computational Intelligence and Neuroscience, 2022*, Article 2938011. https://doi.org/10.1155/2022/2938011

Zhang, H., Venkatramani, S., Paz, D., Li, Q., Xiang, H., & Christensen, H. I. (2023). Probabilistic semantic mapping for autonomous driving in urban environments. *Sensors, 23*(14), Article 6504. https://doi.org/10.3390/s23146504

Zhou, H., Xu, L., Ren, Y., Zhang, D., Li, P., Yang, J., Yan, J., & Tan, Z. (2025). Research on vehicle accident hazard scenario derivation based on improved AST. *Scientific Reports, 15*(1), Article 26350. https://doi.org/10.1038/s41598-025-11593-8

Appendix A

Analysis Method

Figure A1

Mathematical Analysis

$$h_{msr} = \alpha_s + \beta_m + \epsilon_{msr}$$

$$\epsilon_{msr} \sim \mathcal{N}(0, \sigma^2)$$

$$h_{msr} \mid \alpha_s, \beta_m, \sigma^2 \sim \mathcal{N}(\alpha_s + \beta_m, \sigma^2)$$

$$\mathcal{L}(\alpha, \beta, \sigma^2 \mid h) = \prod_{m=1}^{2} \prod_{s=1}^{S} \prod_{r=1}^{R} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(h_{msr} - \alpha_s - \beta_m)^2}{2\sigma^2} \right)$$

$$\alpha_s \sim \mathcal{N}(0, \tau_\alpha^2)$$

$$\beta_m \sim \mathcal{N}(0, \tau_\beta^2)$$

$$\sigma^2 \sim \text{Inverse-Gamma}(a_\sigma, b_\sigma)$$

$$\tau_\alpha^2 \sim \text{Inverse-Gamma}(a_\alpha, b_\alpha)$$

$$\tau_\beta^2 \sim \text{Inverse-Gamma}(a_\beta, b_\beta)$$

$$P(\alpha, \beta, \sigma^2, \tau_\alpha^2, \tau_\beta^2 \mid h) \propto \mathcal{L}(h \mid \alpha, \beta, \sigma^2) P(\alpha \mid \tau_\alpha^2) P(\beta \mid \tau_\beta^2) P(\sigma^2) P(\tau_\alpha^2) P(\tau_\beta^2)$$

$$\alpha_s \mid - \sim \mathcal{N}\left(\frac{\sum_{m,r}(h_{msr} - \beta_m)}{MR + \frac{\sigma^2}{\tau_\alpha^2}}, \frac{\sigma^2}{MR + \frac{\sigma^2}{\tau_\alpha^2}} \right)$$

$$\beta_m \mid - \sim \mathcal{N}\left(\frac{\sum_{s,r}(h_{msr} - \alpha_s)}{SR + \frac{\sigma^2}{\tau_\beta^2}}, \frac{\sigma^2}{SR + \frac{\sigma^2}{\tau_\beta^2}} \right)$$

$$\sigma^2 \mid - \sim \text{Inverse-Gamma}\left(a_\sigma + \frac{N}{2},\ b_\sigma + \frac{1}{2}\sum_{m,s,r}(h_{msr} - \alpha_s - \beta_m)^2\right)$$

$$N = 2SR$$

$$\hat{\beta}_m = E[\beta_m \mid h]$$

$$\Delta = \beta_{\text{Ensemble}} - \beta_{\text{Carla}}$$

$$P(\Delta < 0 \mid h)$$

$$\Delta \in [\Delta_{0.005}, \Delta_{0.995}]$$

$$\text{Var}(h_{msr}) = \tau_\alpha^2 + \tau_\beta^2 + \sigma^2$$

$$\text{ICC}_{\text{map}} = \frac{\tau_\alpha^2}{\tau_\alpha^2 + \tau_\beta^2 + \sigma^2}$$

$$\text{Percent Reduction} = \frac{\bar{h}_{\text{Carla}} - \bar{h}_{\text{Ensemble}}}{\bar{h}_{\text{Carla}}} \times 100$$

$$h_{\text{new}} \sim \mathcal{N}(\alpha_s + \beta_m, \sigma^2)$$

$$P(h_{\text{new}} \mid h) = \int P(h_{\text{new}} \mid \theta) P(\theta \mid h)\, d\theta$$

$$\theta = (\alpha, \beta, \sigma^2)$$

Figure A2

Code for Bayesian Hierarchical Model

```
import numpy as np
import pymc as pm
import arviz as az
ensemble = np.array([
    [1400,  900, 1300, 1200, 1000],
    [2100, 1500, 1900, 1700, 1800],
    [1200,  700, 1000,  900,  800],
    [6500, 5400, 6200, 5900, 5700],
    [2800, 2300, 2600, 2500, 2400],
    [7600, 6500, 7200, 6900, 7000],
    [1300,  700, 1100,  900,  800],
    [8800, 7600, 8300, 8000, 7900],
])
carla = np.array([
    [1500, 1100, 1400, 1300, 1200],
    [2200, 1900, 2100, 2000, 2000],
    [1300,  900, 1200, 1100, 1000],
    [6900, 6100, 6600, 6400, 6200],
    [3000, 2600, 2900, 2700, 2600],
    [8100, 7200, 7800, 7600, 7400],
    [1500,  900, 1300, 1100, 1000],
    [9300, 8600, 9000, 8800, 8700],
])
y = np.concatenate([ensemble.flatten(), carla.flatten()])
vehicle = np.concatenate([
    np.zeros(40),
    np.ones(40)
])
map_idx = np.tile(np.repeat(np.arange(8), 5), 2)
with pm.Model() as model:
    alpha_global = pm.Normal("alpha_global", mu=4000, sigma=3000)
    beta_global  = pm.Normal("beta_global", mu=0, sigma=2000)
    sigma_alpha = pm.HalfNormal("sigma_alpha", sigma=2000)
    sigma_beta  = pm.HalfNormal("sigma_beta", sigma=1500)
    alpha_map = pm.Normal("alpha_map", mu=alpha_global, sigma=sigma_alpha, shape=8)
    beta_map  = pm.Normal("beta_map",  mu=beta_global,  sigma=sigma_beta,  shape=8)
    mu = alpha_map[map_idx] + beta_map[map_idx] * vehicle
    sigma = pm.HalfNormal("sigma", sigma=1000)
    y_obs = pm.Normal("y_obs", mu=mu, sigma=sigma, observed=y)
    t = pm.sample(2000, tune=2000, target_accept=0.9, random_seed=67, progressbar=False)
b = t.posterior["beta_global"].values.flatten()
eb = np.mean(b > 0)
cb = np.mean(b < 0)
d = b.mean()
print(f"Ensemble better {eb:.3f}")
print(f"CARLA better {cb:.3f}")
print(f"damage difference {d:,.0f}")
```

Appendix B

Collected Data

Figure B1

Sample values

	My Model	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Map 1		1400	900	1300	1200	1000
Map 2		2100	1500	1900	1700	1800
Map 3		1200	700	1000	900	800
Map 4		6500	5400	6200	5900	5700
Map 5		2800	2300	2600	2500	2400
Map 6		7600	6500	7200	6900	7000
Map 7		1300	700	1100	900	800
Map 8		8800	7600	8300	8000	7900

	CARLA 2024 Model	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Map 1		1500	1100	1400	1300	1200
Map 2		2200	1900	2100	2000	2000
Map 3		1300	900	1200	1100	1000
Map 4		6900	6100	6600	6400	6200
Map 5		3000	2600	2900	2700	2600
Map 6		8100	7200	7800	7600	7400
Map 7		1500	900	1300	1100	1000
Map 8		9300	8600	9000	8800	8700

Figure B2

Graph of percent differences