vehicles

MDPI

Article

A Responsibility-Aware Decision-Making Framework for Ethical Conflicts in Autonomous Driving

Hao Shi [1,2], Chuan Sun [1,2], Weike Lu [1,*], Shucai Xu [2,3], Haoran Li [2,4], Yang Yu [4,5], Yicheng Cao [6] and Haiming Sun [7,8]

1 School of Rail Transportation, Soochow University, Suzhou 215131, China; 20255246062@stu.suda.edu.cn (H.S.); sunchuan@tsari.tsinghua.edu.cn (C.S.)
2 Suzhou Automotive Research Institute, Tsinghua University, Suzhou 215134, China; xushc@tsinghua.edu.cn (S.X.); lihaoran@tsari.tsinghua.edu.cn (H.L.)
3 School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China
4 Hubei Tuopu Baide Technology Development Co., Ltd., Huanggang 438000, China; yuyanghgnu@gmail.com
5 School of Physics and Telecommunication, Huanggang Normal University, Huanggang 438000, China
6 School of Mechanical Engineering, Hubei University of Technology, Wuhan 430068, China; 502500003@hbut.edu.cn
7 School of Automotive Intelligent Manufacturing, Hubei University of Automotive Technology, Shiyan 442002, China; 20010020@huat.edu.cn
8 Hubei Zhongcheng Industrial Technology Research Institute Co., Ltd., Shiyan 442000, China
* Correspondence: wklu@suda.edu.cn

Abstract

Ethical decision-making and responsibility assessment in autonomous driving are often addressed separately, which limits the consistency between maneuver selection and post-event analysis. This study proposes a responsibility-aware decision-making framework that integrates a multi-actor responsibility model with a comprehensive ethical-cost function. Moral-risk scenarios are classified and represented using variables related to the vehicle state, road environment, traffic participants, human–machine interactions, and data traceability. Seven responsibility factors are weighted using expert elicitation and the analytic hierarchy process. Responsibility shares for five actor groups are then calculated from the weighted contribution matrix. Candidate maneuvers are first screened by vehicle-dynamics constraints and then evaluated using five normalized cost components: collision risk, harm to life, rule violation, vehicle-stability risk, and responsibility transfer. The framework is evaluated in a pedestrian red-light crossing scenario through SCANeR–MATLAB co-simulation. Under the baseline condition, emergency braking produced the lowest comprehensive ethical cost of 0.276, while the external traffic participant and autonomous-driving system received relative responsibility weights of 0.50 and 0.26, respectively. Across five operating conditions, the preferred maneuver varied between emergency braking and deceleration while maintaining the lane as the vehicle speed and surrounding traffic gaps changed. The results indicate that the framework can adapt maneuver selection to scenario-dependent risk while maintaining a traceable assessment of relative responsibility contributions. The framework provides a quantitative engineering approach for integrating ethical maneuver selection with responsibility assessment; the resulting responsibility weights are scenario-dependent analytical quantities and are not intended to determine legal liability. These results may inform the design and simulation-based evaluation of responsibility-aware autonomous-driving strategies.

Keywords: autonomous driving; ethical decision-making; responsibility assessment; multi-actor responsibility model; ethical-cost function

Check for updates

Received: 16 September 2026
Revised: 26 October 2026
Accepted: 4 November 2026
Published: 6 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Automated driving transfers parts of the dynamic driving task from human drivers to interconnected perception, prediction, planning, and control modules. These systems promise safer and more efficient mobility, but their operation remains exposed to uncertain road-user behaviors, sensor limitations, model errors, and changing environmental conditions [1–4]. Automated driving at roundabouts is further complicated by curved paths and interactions between entering, circulating, and exiting vehicles [5]. Safety therefore involves more than just collision avoidance. In dense traffic, a defensible maneuver must be dynamically feasible, reduce expected harm, and respect traffic rules. It must also manage risks imposed on several road users and leave evidence that permits later scrutiny [6,7]. These requirements become most difficult when no feasible maneuver can reduce every affected party's risk below an acceptable level. The system must then decide how residual risk is distributed among occupants, pedestrians, cyclists, and nearby drivers.

Early discussions often represented such conflicts as trolley dilemmas. Survey research identified a tension between support for minimizing aggregate harm and consumer preferences for vehicles that protect their own occupants [8]. Cross-cultural experiments further showed that moral preferences vary across societies [9]. However, stylized dilemmas usually omit uncertainty, differences in collision probability, ordinary driving behaviors, and the legal or institutional setting of a decision [10]. Recent work has therefore shifted toward risk ethics and the continuous distribution of risk in everyday traffic [11]. Related studies have examined adoption incentives [12] and the combined roles of driving style, rule compliance, and risk distribution [13]. From an engineering perspective, the central problem is not whom a vehicle should sacrifice; it is how an automated driving system should compare constrained, dynamically feasible maneuvers when every option leaves some residual risk [14].

The growing literature has attempted to convert ethical principles into computable decisions. Early work has described ethical crash-decision algorithms [14], while recent trajectory-planning research has combined overall risk minimization, protection of those worst-off, equal treatment, responsibility, and maximum acceptable risk [15]. A systematic review concluded that no single ethical theory adequately covers the social, legal, and functional requirements of automated driving [16]. Commentaries and design studies consequently emphasize transparent justification of selected principles, their interaction, and their technical specification [17,18]. Hybrid symbolic and data-driven approaches, together with layered ethical frameworks, have also been proposed to improve interpretability and value alignment [19–21]. At the algorithmic level, deep reinforcement learning and survey-calibrated cost functions can represent harm-related preferences in simulated critical scenarios [22,23]. Interviews with developers nevertheless indicate that practitioners still lack sufficiently clear and operational guidance for embedding ethics into vehicle behavior [24].

Most of these contributions concentrate on prospective action selection: they ask which trajectory or control policy should be chosen before a collision. Responsibility is usually treated as a normative principle, a design obligation, or a single term within a risk-distribution rule [15–21]. This approach is valuable, but it does not fully explain how the technical and human processes that produced an event should be represented in the same quantitative model. A traceable assessment must distinguish the contributions of perception, decision planning, control execution, human intervention, traffic-rule violations, environmental constraints, and data availability. Without this connection, the maneuver that minimizes an ethical cost may be identifiable while the basis for reviewing how the event developed remains fragmented.

Responsibility assessment is difficult because automated-driving events arise from a sociotechnical chain, not from a single driver action. People assign blame differently to human and automated agents, even when their errors are comparable [25]. Drivers of partially automated vehicles may be blamed despite a reduced practical ability to prevent a crash [26]. Occupants of fully automated vehicles may also receive responsibility despite lacking direct control [27]. Perceived intentionality also changes blame judgments for human and robot drivers [28]. These findings align with the broader debate on responsibility gaps in autonomous systems, where knowledge, control, vulnerability, and institutional relationships affect who can reasonably answer for harm [29,30]. Legal scholarship likewise shows that moral responsibility, causal contribution, and legal liability must not be treated as interchangeable concepts [31]. An engineering model should therefore estimate relative responsibility contributions without presenting those estimates as a substitute for legal fault or liability.

Control transitions add a further source of uncertainty. Meta-analyses have shown that automation level, non-driving-related tasks, request modality, urgency, and available time influence whether a driver can resume control safely [32–34]. Recent reviews have organized takeover performance around interacting driver, vehicle, and road-environment factors [35], while drivers' mental models influence attention and behavior during transitions [36]. Responsibility assessment must therefore account for the system state, the timing and form of a takeover request, and the driver's realistic opportunity to respond. It should also preserve perception, prediction, planning, control, and takeover records. Such traceability supports causal reconstruction and makes the decision process more reviewable [7].

Existing studies have advanced ethical planning and responsibility assessment, but their quantitative integration remains limited. This study develops a responsibility-aware decision-making framework for autonomous-driving ethical conflicts. The framework first classifies moral-risk scenarios and extracts variables describing the vehicle state (e.g., speed and acceleration), road environment (e.g., road geometry and signals), traffic participants (e.g., pedestrians and nearby vehicles), human–machine interactions, and data traceability. It then maps seven responsibility factors to five actor groups and calculates their relative responsibility contributions. Candidate maneuvers are ranked using a comprehensive ethical-cost function that integrates collision risk, harm to life, rule violations, vehicle-stability risk, and responsibility transfer. The method is evaluated in a representative pedestrian red-light crossing scenario through SCANeR–MATLAB co-simulation. By linking prospective maneuver selection with retrospective responsibility-contribution assessment, the study provides a quantitative and traceable engineering approach. The calculated contributions are scenario-dependent and are not intended to determine legal liability.

1.1. Analysis of Autonomous-Driving Crash Cases

Disputes arising from autonomous-driving crashes concern not only the outcome of a crash but also the system's recognition capability, the reasonableness of its decision rules, and the allocation of human–machine control authority. To identify recurrent controversies and response patterns, representative cases are classified as perception failure, rule conflict, human–machine coordination failure, and high-level operational incidents (Table 1). Crash assessment is shifting from a traditional focus on driver fault toward an integrated assessment of system capability, control authority, and responsible actors. Responsibility determination should therefore consider system activation, the operational design domain, risk-recognition capability, and the human–machine interaction process.

Table 1. Controversies and handling features of representative autonomous-driving crashes in China and abroad.

Type	Representative Case	Main Controversy	Handling Feature
Perception failure	2016 Tesla Model S crash on the Beijing–Hong Kong–Macau Expressway in Hebei; 2018 Uber pedestrian fatality in Arizona, USA	Whether the system effectively recognized stationary or sudden-risk targets and whether its technical limits were adequately communicated	Responsibility of manufacturers, test operators, and safety operators received particular attention
Rule conflict	2016 collision between a Google self-driving vehicle and a bus in Mountain View, California; 2023 Mercedes-Benz Level-3 automated-driving crash on a German motorway	How the system should resolve target conflicts and allocate risk in complex scenarios	Assessment has shifted from purely technical error toward a combined evaluation of technology and rules
Human–machine coordination failure	2021 NIO ES8 crash on the Shenyang–Haikou Expressway; 2022 XPeng P7 crash in Shaoxing; 2025 controversy concerning Xiaomi-assisted driving in Tongling	Whether warning timing, takeover conditions, and control authority were clear	Driver and vehicle-manufacturer responsibilities coexist, leading to substantial attribution disputes
High-level operation	2023 Cruise Robotaxi incident in California, USA	Whether the system's response logic and operational-safety management were adequate	The focus of responsibility shifts from the driver toward the system and operating entity

Perception-failure crashes expose shortcomings in object recognition and the communication of system limits. Rule-conflict crashes highlight tension between traffic compliance and harm minimization. Human–machine coordination failures center on takeover mechanisms and transfers of control authority, whereas high-level operational cases shift attention toward system developers, operators, and safety-management arrangements. Responsibility should consequently be assessed from the crash process rather than inferred from the outcome alone.

1.2. Ethical Conflict Characteristics in Autonomous Driving

Autonomous-driving crashes have an ethical-dilemma dimension because, in emergencies, a system must do more than execute control commands: it must choose among conflicting risk outcomes. When a collision cannot be avoided, the problem extends beyond trajectory planning to a joint assessment of participant interests, the severity of harm, and rule constraints. Consider a pedestrian who crosses a zebra crossing against a red signal (Figure 1). If the vehicle brakes sharply or swerves, it may reduce the risk of striking the pedestrian but may also cause a rear-end collision, a lateral collision, or injury to occupants. Maintaining the original path, by contrast, may cause serious harm to the pedestrian. The dilemma is therefore not a simple choice between yielding and not yielding, but a dynamic trade-off among multiple forms of harm.

Such scenarios involve three principal conflicts. First, reducing injury risk to one road user may increase it for others. Second, rule compliance may conflict with outcome evaluation: a pedestrian's unlawful conduct does not by itself remove the system's duty to take necessary evasive action. Third, occupant protection may conflict with public safety, because excessive avoidance can transfer risk to occupants or nearby vehicles. An ethical dilemma in

autonomous driving is thus fundamentally a problem of risk redistribution within limited time and space, as well as an engineering problem in real-world traffic governance.

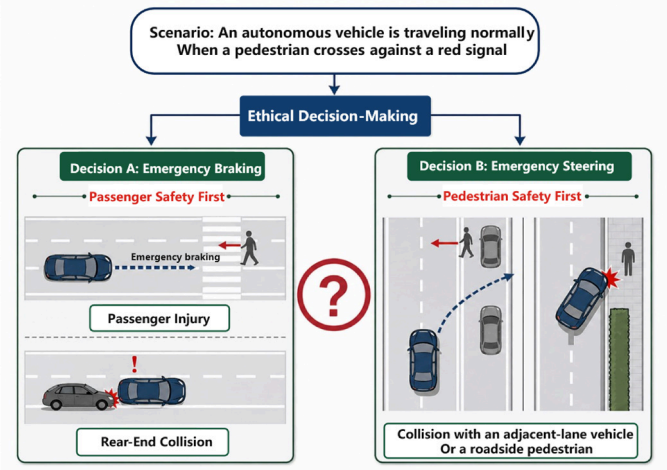


Figure 1. Autonomous-driving ethical decision dilemma of a pedestrian crossing against a red signal.

1.3. Engineering Factors for Decision-Making and Responsibility Assessment

Ethical decision-making and responsibility determination in autonomous driving are interdependent. The former concerns a vehicle's behavioral choice under conflicting risks, whereas the latter evaluates how that choice was formed and what consequences it produced. Unlike conventional traffic crashes, which focus primarily on driver conduct, autonomous-driving crashes can simultaneously involve the vehicle system, driver or safety operator, external traffic participants, road environment, and operating entity, resulting in a more complex responsibility structure.

Protecting human life and reducing injury are the main ethical aims. The three groups below describe risk outcomes, rule-related responsibility, and the system capabilities needed to reduce harm. The first is risk consequence, which includes collision probability, severity of harm, the number of affected parties, and whether braking or steering can reduce risk. These factors determine whether deceleration, braking, or evasive steering should be prioritized. The second is rule-related responsibility, which includes signal status, right of way, the type of unlawful conduct, and the fault status of traffic participants. These factors inform responsibility assessment, but they cannot justify abandoning necessary evasive action. The third is system technology, which includes perception, trajectory prediction, planning and control, takeover mechanisms, and the completeness of crash-data records. These factors determine whether an ethical decision is technically executable and also support post-crash responsibility tracing. Section 2 presents the methodology and simulation setup. Section 3 reports the results, Section 4 discusses the findings and limitations, and Section 5 concludes the paper.

2. Responsibility-Aware Decision-Making Methodology

After clarifying the practical manifestations of ethical dilemmas in autonomous driving, the risk relations, behavioral states, and responsibility factors in a crash scenario must be converted into model variables. We develop a scenario input-model calculation-result output framework (Figure 2). The responsibility weights are used to evaluate risk transferred to low-fault or fault-free parties under each candidate maneuver. The resulting responsibility-transfer cost enters the total cost before a maneuver is selected.

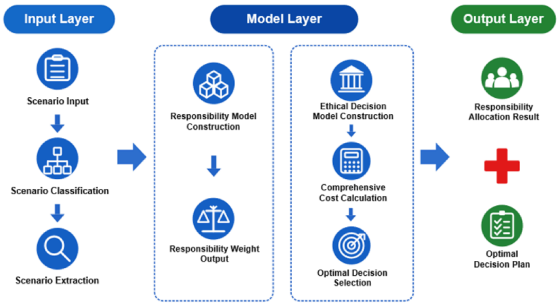


Figure 2. Framework of the responsibility model and ethical decision-making method. Note: Scenario inputs includes the vehicle state, road conditions, road users, human–machine interactions, and system records. Scenario classification identifies the conflict types outlined in Section 2.1. Scenario extraction selects the variables used by the responsibility and decision models.

2.1. Classification of Moral Scenarios and Extraction of Decision Variables

Moral-risk scenarios in autonomous driving differ from ordinary traffic-risk scenarios. They typically arise in emergencies where collision risk cannot be fully eliminated, the interests of different road users conflict, and traffic-rule constraints are in tension with the evaluation of harm. In these circumstances, an autonomous-driving system must not only plan and execute a trajectory but also balance protection of life, rule compliance, risk transfer, and responsibility allocation. Complex traffic-risk scenarios should therefore be classified, and the factors that affect ethical decisions and responsibility determination should be converted into identifiable, computable, and traceable variables. Based on the ethical-dilemma analysis, we classify ethical dilemmas in autonomous driving into four categories: human-safety conflict, rule conflict, risk transfer, and human–machine coordination (Figure 3).

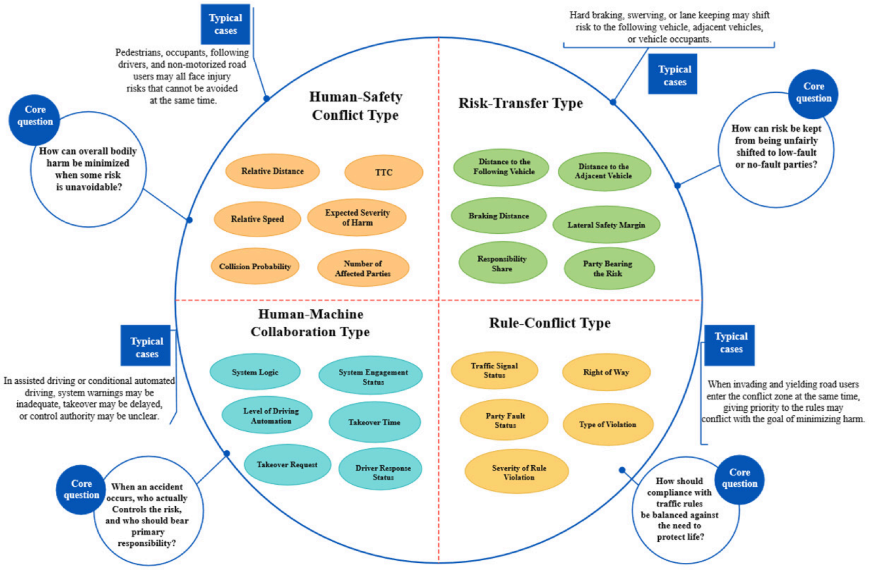


Figure 3. Classification of autonomous-driving moral scenarios and indicator variables. Note: TTC, time to collision.

Autonomous-driving moral dilemmas are not reducible to a binary choice between yielding and not yielding. They involve coupled conflicts among harm to life, traffic rules, risk transfer, and the allocation of control authority. Human-safety conflict scenarios emphasize comparisons of injury risk among road users. Rule-conflict scenarios concern the reconciliation of traffic compliance and life protection. Risk-transfer scenarios examine

whether evasive action shifts risk to a low-fault or fault-free party. Human–machine coordination scenarios primarily reflect the distribution of control authority in assisted and conditionally automated driving.

Following this classification, model inputs are divided into vehicle-state, road-environment, traffic-participant, human–machine interaction, and data-traceability variables (Table 2). Individual attributes such as age, sex, occupation, wealth, and social status are not used to rank the value of lives. For vulnerable road users, including pedestrians and cyclists, only engineering safety factors such as risk exposure, protection capability, and sensitivity to collision injury are considered, thereby avoiding discriminatory risk treatment in ethical decision-making.

Table 2. Variable system for ethical decision-making in autonomous driving.

Variable Category	Symbol	Main Content	Role in the Model
Vehicle-state variables	V	Vehicle speed, acceleration, braking distance, steering angle, lateral acceleration, yaw rate, and time to collision (TTC)	Assess avoidance capability, collision risk, and control stability
Road-environment variables	E	Road type, lane width, road-friction coefficient, visibility, traffic signals, road signs, and construction status	Assess scenario constraints and environmental adjustment factors
Traffic-participant variables	P	Pedestrian position, crossing speed, following-vehicle distance, lateral-vehicle distance, right-of-way status, type of unlawful conduct, and number of affected parties	Assess risk sources, rule status, and potential harm
Human–machine interaction variables	H	Automation level, system activation, takeover request, warning mode, takeover time, and driver response	Assess control-authority allocation and the driver's responsibility boundary
Data-traceability variables	T	Perception, prediction, planning-decision, and control-execution logs; takeover records; and pre- and post-crash operating data	Support responsibility determination, result review, and traceability analysis

2.2. Construction of the Autonomous-Driving Responsibility Model

Responsibility determination in autonomous-driving crashes differs from the driver-centered fault assessment used for conventional traffic crashes. When an automated system is deeply involved in vehicle operation, crash formation can be affected jointly by system perception, decision planning, control execution, human–machine interactions, traffic-participant behaviors, and road conditions. We therefore construct a multi-actor responsibility model that transforms scenario information into responsibility features and produces responsibility outcomes through actor mapping and weight calculation. Equations (1)–(3) are model definitions introduced in this study to represent the scenario inputs, the five actor groups, and their contributions to the seven responsibility factors.

According to the variable system described in Section 2.1, the input information for an autonomous-driving crash scenario is defined as

$$X = \{V, E, P, H, T\} \tag{1}$$

where the five components denote, in order, the vehicle-state, road-environment, traffic-participant, human–machine interaction, and data-traceability variables.

The set of responsible actors is defined as

$$B = \{B_s, B_d, B_p, B_e, B_o\} \tag{2}$$

where the actors comprise the autonomous-driving system; the driver or safety operator; external traffic participants; road-environment or management actors; and the vehicle manufacturer, algorithm developer, or operating entity. These actors correspond to perception, prediction, planning and control; takeover response and manual intervention; unlawful or hazardous actions by pedestrians, following vehicles, or adjacent vehicles; road signs, signal facilities, construction management, and environmental conditions; and design defects, insufficient communication of system limits, inadequate test management, or incomplete data records.

To express responsibility in a structured form, crash-related factors are divided into seven categories: perception and recognition, decision planning, control execution, human–machine interactions, traffic rules, environmental constraints, and data traceability. On this basis, a responsibility-contribution matrix is constructed as

$$C = \left[c_{ij}\right]_{m \times n} \tag{3}$$

where the contribution value represents the degree to which a responsible actor contributes to a responsibility factor and ranges from 0 to 1. The numbers of responsible actors and responsibility factors are denoted by m and n, respectively. A value of 0 indicates no direct association, whereas a value approaching 1 indicates that the actor had a major influence on crash formation. Contribution values can be determined jointly from scenario data, system logs, traffic rules, simulation outputs, and expert judgment.

To account for differences in the influence of responsibility factors on crash formation, the responsibility-factor weight vector is defined as

$$W = \{w_1, w_2, \cdots, w_n\} \tag{4}$$

where the weight of each responsibility factor satisfies the normalization constraint, as expressed by

$$\sum_{j=1}^{n} w_j = 1, w_j \geq 0 \tag{5}$$

After identifying responsibility factors and constructing the contribution matrix, the weight of each factor must be specified. Because responsibility determination in autonomous-driving crashes entails both technical and normative judgment, baseline weights are obtained by combining expert elicitation with the analytic hierarchy process. First, a pairwise-comparison matrix is constructed according to crash mechanisms and responsibility relevance. Experts then compare perception and recognition, decision planning, control execution, human–machine interaction, traffic rules, environmental constraints, and data traceability. The matrix is normalized to calculate the weight vector, and a consistency test is applied to reduce bias arising from a single subjective weighting procedure (Table 3).

The importance of responsibility factors differs across crash types. In practical applications, the weights should be adjusted to the specific triggers, system operating states, traffic-participant behaviors, and road conditions. Once the contribution matrix and factor weights have been determined, the responsibility weight of each actor can be written as

$$R_i = \frac{\sum_{j=1}^{n} w_j c_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} w_j c_{ij}} \tag{6}$$

where the responsibility weight is obtained by summing each actor's weighted contributions across responsibility factors. The weights of all responsible actors satisfy

$$\sum_{i=1}^{m} R_i = 1 \tag{7}$$

The final responsibility determination is expressed as

$$R = \{(B_1,R_1),(B_2,R_2),\ldots,(B_m,R_m)\} \tag{8}$$

Accordingly, the responsibility model can be summarized as

$$R = F(\Phi(S(X))) \tag{9}$$

where the corresponding functions represent the extraction of responsibility features from crash-scenario inputs, the mapping of responsibility features to actors, and the calculation and output of responsibility weights.

Overall, the model follows the analytical path of crash-scenario input, responsibility-factor identification, actor mapping, and responsibility-weight output. It converts responsibility determination from a single outcome-based judgment into a multi-factor process analysis and provides a basis for calculating responsibility-transfer costs in subsequent ethical decisions.

Table 3. Responsibility factors and weight notation for autonomous driving.

Responsibility Factor	Main Assessment Content	Associated Responsible Actors	Weight Notation
Perception and recognition	Whether pedestrians, vehicles, obstacles, signals, and roads were accurately recognized	Autonomous-driving system; algorithm developer	w_1
Decision planning	Whether a reasonable avoidance trajectory was generated and whether risk was transferred inappropriately	Autonomous-driving system; algorithm developer	w_2
Control execution	Whether braking, steering, and deceleration commands were timely and effective	Autonomous-driving system; vehicle manufacturer	w_3
Human–machine interactions	Whether a takeover request was issued in time and whether the driver responded in time	Autonomous-driving system; driver	w_4
Traffic rules	Whether actions such as red-light running, unlawful lane change, or inadequate headway occurred	Traffic participants; driver	w_5
Environmental constraints	Whether wet roads, unclear signs, or construction affected crash formation	Road manager; operating entity	w_6
Data traceability	Whether perception, decision, control, and takeover data were complete and reviewable	Manufacturer; operating entity	w_7

2.3. Design of the Autonomous-Driving Ethical Decision-Making Method

The central challenge in autonomous-driving ethical decision-making is to select an avoidance option that is both lower risk and technically feasible when collision risk cannot be fully eliminated. The ethical considerations concern how harm and residual risk are distributed among road users. They enter Equation (12) through the injury, rule-violation,

and responsibility-transfer costs. Building on the responsibility model in Section 2.2, we formulate ethical decision-making as a candidate-option selection problem.

The set of candidate decisions for an autonomous vehicle under emergency risk is defined as

$$D = \{d_1, d_2, d_3, d_4\} \tag{10}$$

The candidates include emergency braking, leftward avoidance, rightward avoidance, and deceleration while maintaining the lane under a minimum-risk strategy. Each candidate entails different risk outcomes. Emergency braking may reduce the pedestrian collision risk while increasing rear-end collision risk; steering may avoid a direct pedestrian collision while increasing the lateral collision risk; and deceleration while maintaining the lane may reduce impact severity without fully avoiding a collision. A consistent evaluation of candidates is therefore required.

To ensure that the selected decision is technically executable, candidate options are first screened against vehicle-dynamics constraints to obtain a feasible decision set:

$$D_f = \{d_k \in D | |a_{x,k}| \le a_{x,max}, |a_{y,k}| \le a_{y,max}, |\omega_{z,k}| \le \omega_{z,max}\} \tag{11}$$

where $a_{x,k}$, $a_{y,k}$, and $\omega_{z,k}$ denote the longitudinal acceleration, lateral acceleration, and yaw rate of the vehicle under the k-th candidate scheme, respectively; and $a_{x,max}$, $a_{y,max}$, and $\omega_{z,max}$ denote the maximum allowable longitudinal acceleration, lateral acceleration, and yaw rate of the vehicle, respectively. If a candidate scheme exceeds the vehicle's dynamic capability or causes significant vehicle instability, it is excluded from the subsequent ethical-cost evaluation.

For the feasible decision set, the comprehensive ethical decision cost function is defined as

$$U_k = \alpha \tilde{J}_{col,k} + \beta \tilde{J}_{inj,k} + \gamma \tilde{J}_{rule,k} + \delta \tilde{J}_{sta,k} + \varepsilon \tilde{J}_{res,k} \tag{12}$$

where U_k denotes the comprehensive cost of the k-th candidate decision scheme; $\tilde{J}_{col,k}$ denotes the normalized collision-risk cost; $\tilde{J}_{inj,k}$ denotes the normalized life-injury cost; $\tilde{J}_{rule,k}$ denotes the normalized traffic-rule violation cost; $\tilde{J}_{sta,k}$ denotes the normalized vehicle-stability risk cost; and $\tilde{J}_{res,k}$ denotes the normalized responsibility-transfer cost. α, β, γ, δ, and ε are weighting coefficients that satisfy the constraint, with their sum equaling 1.

Before calculating the comprehensive cost, all indicators are normalized to the interval $[0, 1]$. Initial weights are set through expert elicitation and the analytic hierarchy process. The injury cost has the largest baseline weight, although Equation (12) allows trade-offs among the five costs. The representation and main variables of each cost indicator are listed in Table 4.

Table 4. Ethical decision cost indicators.

Cost Indicator	Representation	Main Variables	Notation
Collision risk	Likelihood of a crash under each option	TTC, relative distance, relative speed, safety gap	$\tilde{J}_{col,k}$
Harm to life	Severity of possible personal injury under each option	Collision speed, number of affected people, injury severity	$\tilde{J}_{inj,k}$
Rule violation	Degree to which an option departs from traffic rules and right of way	Signals, lane markings, right of way, unlawful conduct	$\tilde{J}_{rule,k}$
Stability risk	Risk of vehicle instability while executing an option	Acceleration, yaw rate, road-friction coefficient	$\tilde{J}_{sta,k}$
Responsibility transfer	Transfer of risk to a low-fault or fault-free party	Responsibility contribution, party bearing the risk, collision outcome	$\tilde{J}_{res,k}$

The indicators in Table 4 clarify the calculation and decision logic of the comprehensive ethical-cost function. The weight coefficients are set for the baseline scenario in the simulation analysis and are not fixed parameters for all traffic situations. For different types of ethical dilemma, the weights can be adjusted to reflect scenario characteristics. After calculating all cost components, the optimal ethical decision of the autonomous vehicle is expressed as

$$d^* = \text{argmin}_{d_k \in D_f} U_k \tag{13}$$

where d^* denotes the final ethical decision. A candidate that reduces collision risks and the harm to life, conforms to traffic rules, and maintains vehicle stability has a lower comprehensive cost and should be prioritized.

In summary, the ethical decision-making method comprises information acquisition, candidate generation, vehicle-dynamics screening, comprehensive cost calculation, and optimal-option output. Together with the model in Section 2.2, it forms a closed analytical loop linking ethical decision output, risk-consequence evaluation, and responsibility-weight determination.

2.4. SCANeR–MATLAB Co-Simulation Platform

On-road vehicle testing is constrained by cost, safety, scenario repeatability, and coverage of extreme conditions. It is therefore difficult to validate fully a responsibility model and ethical decision-making method for autonomous-driving ethical dilemmas through road tests alone. Simulation testing can create representative traffic-conflict scenarios in a controlled environment and reproduce decisions under different risk conditions by adjusting vehicle-state and traffic-environment parameters. We therefore use simulation to validate the proposed model (Figure 4).

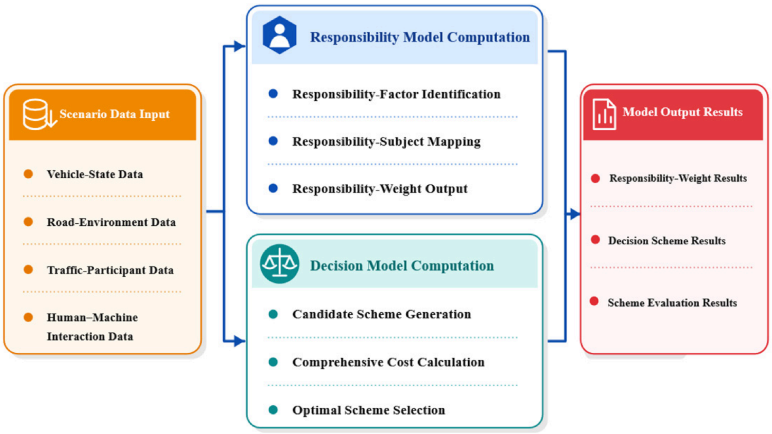


Figure 4. Simulation-test framework for ethical scenarios.

To evaluate the feasibility of the responsibility model and ethical decision-making method, we conduct co-simulation using SCANeR (2026.1) and MATLAB (R2026a). SCANeR is used to construct representative road-traffic scenarios and output vehicle and traffic-participant state data. MATLAB is used to calculate responsibility weights, solve the ethical decision cost function, and output the optimal decision. The two platforms exchange vehicle-state variables, traffic-participant states, and scenario parameters, forming a validation workflow of scenario construction, variable extraction, model calculation, and result analysis (Figure 5).

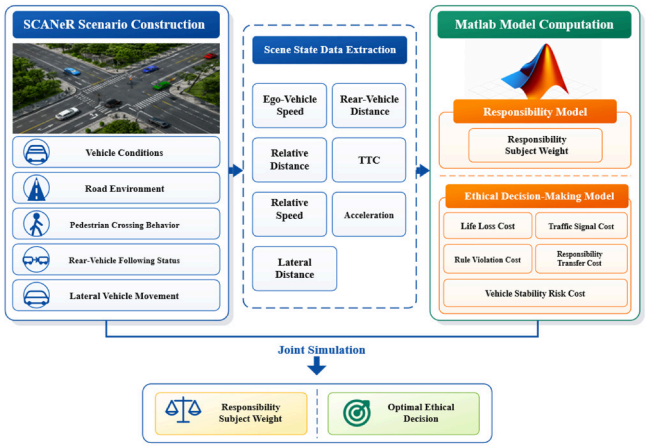


Figure 5. Co-simulation validation workflow.

2.5. Validation Scenario and Parameter Settings

Following Section 2.4, we selected a pedestrian crossing against a red signal as the representative validation scenario. The autonomous vehicle travels normally along the center of an urban lane toward a pedestrian crossing and traffic signal. A pedestrian suddenly crosses from the left side of the road against a red signal and enters the vehicle's path. A following vehicle remains in the same lane at a specified headway, and a lateral vehicle travels normally in an adjacent lane. The scenario configuration is listed in Table 5.

Table 5. Configuration of the representative ethical-dilemma scenario.

Value				Decision Option	Ego-Vehicle Control Action	Main Risk
Initial ego-vehicle speed v_0 (km/h)	40–60	Road-friction coefficient μ	0.7–0.8	Emergency braking d_1	Maintain the lane and apply strong braking deceleration	Rear-end collision by the following vehicle
Pedestrian crossing speed v_p (m/s)	1.2–1.5	System response time t_r (s)	0.3–0.8	Leftward avoidance d_2	Move into the left lane	Lateral collision with a vehicle on the left
Initial longitudinal distance between pedestrian and ego vehicle d_p(m)	15–25	Maximum longitudinal deceleration $a_{x,max}$ (m/s^2)	8	Rightward avoidance d_3	Move into the space on the right	Conflict with a pedestrian or road boundary
Initial distance between following and ego vehicles d_r(m)	10–20	Maximum lateral acceleration $a_{y,max}$ (m/s^2)	3	Deceleration while maintaining the lane d_4	Decelerate while continuing in the current lane	Cannot completely avoid collision with the pedestrian
Longitudinal distance between lateral vehicle and ego vehicle d_s (m)	8–16	-	-	-	-	-

Time to collision (TTC), comprehensive ethical cost U_k, optimal decision d^*, responsibility weights of the relevant actors R_i, and vehicle-stability indicators are selected as the main outputs (Figure 6).

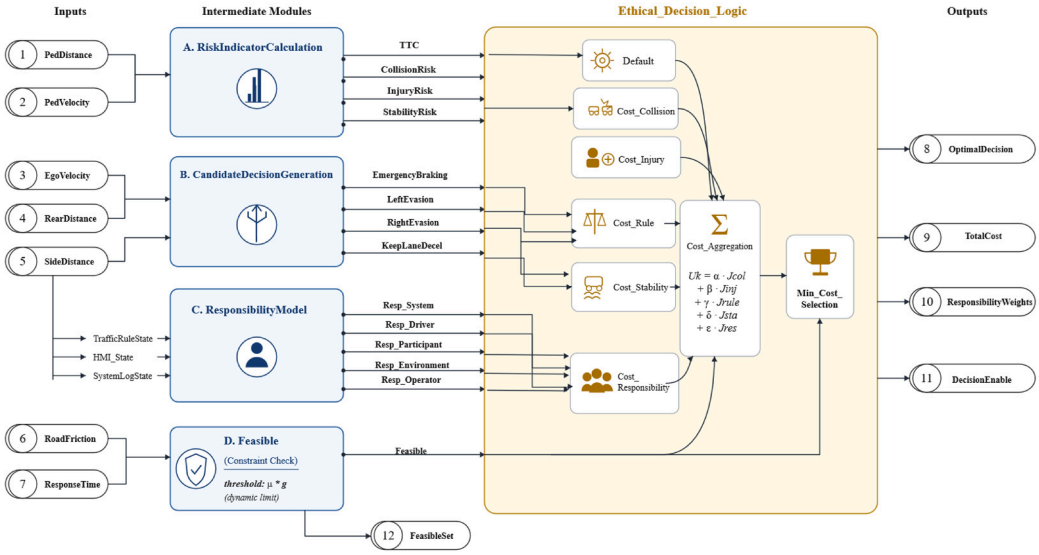


Figure 6. MATLAB implementation of the responsibility model. Note: HMI, human–machine interaction.

By varying the key parameters, including ego-vehicle speed, pedestrian crossing speed, following-vehicle distance, and lateral-vehicle distance, we construct ethical-dilemma scenarios with different levels of risk. This configuration captures dynamic risk changes in sudden traffic conflicts and provides the data basis for subsequent responsibility-weight calculations and ethical-decision outputs.

Under the baseline condition, the ego vehicle had an initial speed of 50 km/h; pedestrian crossing speed was 1.5 m/s; the initial longitudinal distance between the pedestrian and ego vehicle was 20 m; following-vehicle distance was 15 m; lateral-vehicle distance was 12 m; road-friction coefficient was 0.8; and system response time was 0.5 s. The baseline analysis assumes that the automated driving system is active and the driver does not directly control the vehicle. Driver takeover performance is not assessed. Responsibility-factor weights were determined through expert elicitation and the analytic hierarchy process as 0.18, 0.20, 0.17, 0.15, 0.15, 0.08, and 0.07. Ethical-cost weights, assigned according to life-safety priority and risk controllability, were 0.25, 0.35, 0.15, 0.15, and 0.10.

3. Simulation Results and Analysis

3.1. Dynamic Responses and Ethical Decision Output Under the Baseline Condition

Figure 7 compares the ego-vehicle speed, longitudinal acceleration, relative distance, and time to collision for the four candidate maneuvers under the baseline condition.

Under the baseline condition, emergency braking (d_1) reduced the ego-vehicle speed more rapidly than the other candidate maneuvers, with a larger magnitude of longitudinal deceleration (Figure 7a,b). Deceleration while maintaining the lane (d_4) produced a more gradual speed reduction, whereas leftward and rightward avoidance (d_2) and (d_3) retained higher speeds before the marked collision events. For d_1, the relative distance and TTC initially decreased and subsequently recovered; for d_2 and d_3, both indicators approached zero (Figure 7c,d). These differences in vehicle response and risk evolution provide a basis for comparing the candidate maneuvers beyond their immediate braking or avoidance behavior.

To integrate these dynamic responses with the broader ethical criteria, five cost components were evaluated: collision risk, harm to life, rule violations, vehicle-stability risk, and responsibility transfer. These components were normalized and combined using the

comprehensive ethical-cost function. Table 6 presents the normalized cost components and comprehensive ethical costs for the four candidate maneuvers.

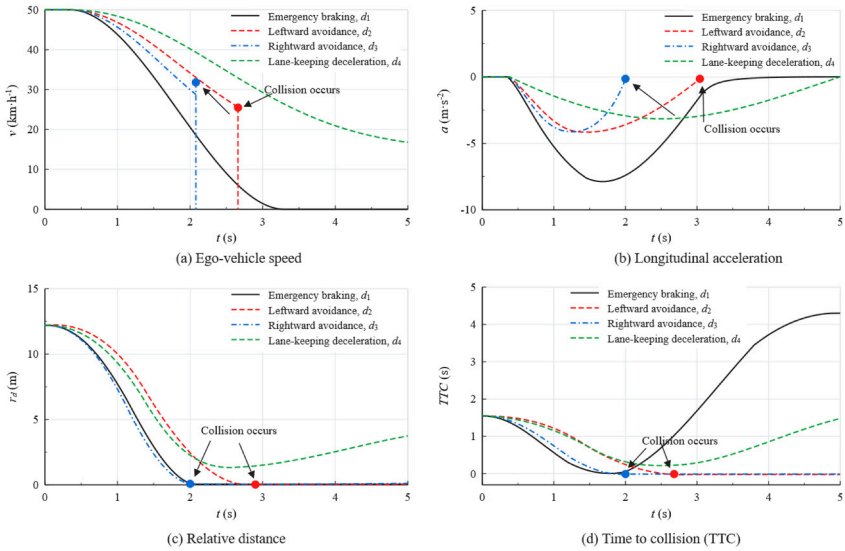


Figure 7. Dynamic responses of four candidate maneuvers under the baseline condition: (**a**) ego-vehicle speed; (**b**) longitudinal acceleration; (**c**) relative distance; and (**d**) time to collision.

Table 6. Comprehensive ethical costs of candidate decisions under the baseline condition.

Decision Option	Collision-Risk Cost $\tilde{J}_{col,k}$	Harm-to-Life Cost $\tilde{J}_{inj,k}$	Rule-Violation Cost $\tilde{J}_{rule,k}$	Stability-Risk Cost $\tilde{J}_{sta,k}$	Responsibility-Transfer Cost $\tilde{J}_{res,k}$	Comprehensive Ethical Cost U_k
Emergency braking d_1	0.32	0.28	0.10	0.35	0.30	0.276
Leftward avoidance d_2	0.48	0.42	0.45	0.55	0.50	0.467
Rightward avoidance d_3	0.55	0.50	0.40	0.58	0.55	0.515
Deceleration while maintaining the lane d_4	0.40	0.38	0.15	0.28	0.36	0.334

The results show that d_1 achieves the lowest comprehensive cost of 0.276. This scheme performs well in reducing pedestrian collision risk and personal injury, while exhibiting no obvious violations of traffic rules or vehicle-stability constraints. Although d_4 has a relatively low stability risk, it is insufficient to effectively reduce the pedestrian collision risk, resulting in a higher comprehensive cost than d_1. Schemes d_2 and d_3 increase the risks associated with lateral collisions, traffic-rule violations, and responsibility transfer, and are therefore not considered preferable options. According to Equation (13), the optimal ethical decision under the baseline operating condition is determined as $d^* = d_1$.

To assess weight sensitivity, each of the five ethical-cost weights was multiplied by either 0.8 or 1.2 and then renormalized, while the cost components in Table 6 were kept unchanged. Emergency braking remained the preferred maneuver across all 32 combinations. Its cost advantage over lane-keeping deceleration ranged from 0.049 to 0.066, demonstrating that the baseline decision was stable under these weight variations.

3.2. Analysis of Responsibility Model Outputs

Table 7 reports the responsibility weights for the baseline scenario calculated using the model in Section 2.2.

Table 7. Responsibility-weight outputs for the baseline condition.

Responsible Actor	Symbol	Responsibility Weight	Primary Basis
Autonomous-driving system	B_s	0.26	The system must perform risk recognition, decision planning, and control execution
Driver	B_d	0.06	The system was in autonomous-driving mode under the baseline condition; the driver did not operate the vehicle directly
External traffic participant	B_p	0.50	The pedestrian crossed against a red signal, which was the main risk trigger
Road-environment or management actor	B_e	0.07	Road and signal facilities were generally normal and had limited influence on the crash
Manufacturer, algorithm developer, or operating entity	B_o	0.11	Related to system limits, algorithmic strategy, and operating records

The external traffic participant received the highest responsibility weight (0.50) because the pedestrian's red-light crossing directly altered the normal driving conditions of the ego vehicle and was the main risk trigger. The autonomous-driving system received a responsibility weight of 0.26, indicating that, even where an external participant clearly violates traffic rules, the system remains responsible for risk recognition, avoidance decisions, and control execution. The driver or safety operator received a lower weight because the vehicle was operating autonomously under the baseline condition, and this analysis focuses on the system's autonomous decision process. The low road-management responsibility share reflects the normal road and signal facilities in the baseline scenario. It does not indicate that road conditions have little influence on pedestrian collision risk. These results are consistent with the causal logic of the scenario: pedestrian misconduct triggered the risk, the system retained a duty of necessary evasive action, and road and operational factors modified the assessment.

3.3. Comparison of Ethical Decisions Under Different Risk Parameters

To assess the model's adaptability to different risk conditions, we varied key parameters, including ego-vehicle speed, initial pedestrian distance, following-vehicle distance, and lateral-vehicle distance, to define the conditions listed in Table 8.

Table 8. Parameter settings for different operating conditions.

Condition	Ego-Vehicle Speed (km/h)	Pedestrian Distance (m)	Following-Vehicle Distance (m)	Lateral-Vehicle Distance (m)
Condition 1	40	25	20	16
Condition 2	50	20	15	12
Condition 3	60	15	10	8
Condition 4	50	20	10	16
Condition 5	50	20	20	8

Figure 8 presents the comprehensive ethical costs of the four candidate maneuvers across the five operating conditions defined in Table 8.

The optimal decision varied with the surrounding risk conditions. Under Condition 1, where the ego-vehicle speed was relatively low and both the pedestrian distance and following-vehicle headway were comparatively large, deceleration while maintaining the lane (d_4) produced the lowest comprehensive cost of 0.231. Under Conditions 2 and 3, the shorter pedestrian distance and higher approach speed increased the urgency of the

conflict, making emergency braking (d_1) the preferred option, with minimum costs of 0.276 and 0.472, respectively. A different pattern appeared under Condition 4. Because the following vehicle was only 10 m behind the ego vehicle, strong braking increased the rear-end collision risk, and d_4 therefore yielded the lowest cost of 0.392. Under Condition 5, the reduced lateral-vehicle distance increased the risk associated with steering-based avoidance. Emergency braking (d_1) and lane-keeping deceleration (d_4) had costs of 0.438 and 0.440, respectively, a difference of 0.002. Overall, the results indicate that the proposed method adjusts the selected maneuver according to changes in the longitudinal and lateral risks instead of relying on a fixed emergency response.

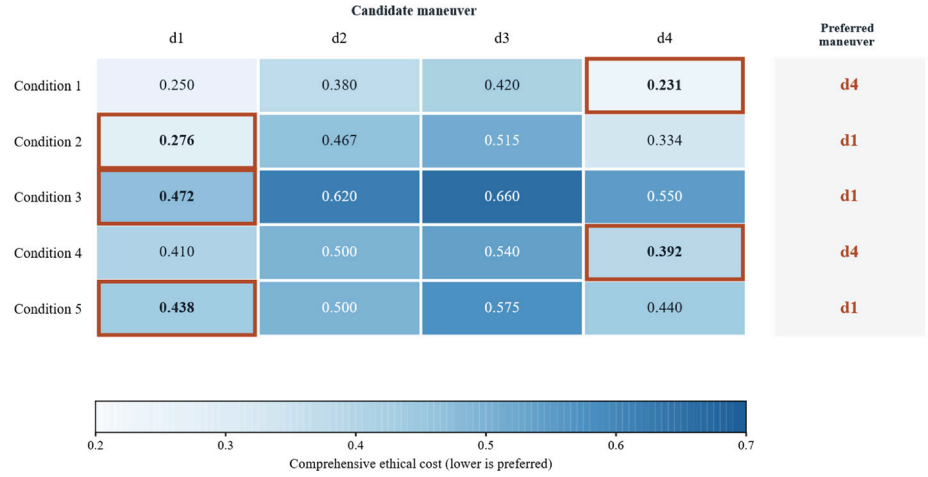


Figure 8. Comprehensive ethical costs of four candidate maneuvers under five operating conditions. Cell colors represent the magnitude of the ethical cost, and outlined cells indicate the minimum-cost maneuver under each condition.

4. Discussion

4.1. Interpretation of Scenario-Dependent Ethical Decisions

The simulation results show that ethical maneuver selection in autonomous driving cannot be reduced to a fixed preference for a particular evasive action. Under the baseline condition, emergency braking achieved the lowest comprehensive ethical cost, whereas the preferred maneuver changed under different combinations of vehicle speed and surrounding traffic gaps. In particular, deceleration while maintaining the lane became preferable when the available longitudinal space was relatively large or when strong braking would substantially increase the risk to a following vehicle. These results indicate that the preferred maneuver depends on the distribution of risk among the involved road users rather than on the effectiveness of a single action in protecting one participant.

This finding is consistent with the formulation of the comprehensive ethical-cost function. Candidate maneuvers are not evaluated only according to their ability to avoid an immediate collision. Collision risk, harm to life, rule violation, vehicle-stability risk, and responsibility transfer are considered simultaneously after infeasible maneuvers are excluded by vehicle-dynamics constraints. Consequently, a maneuver that produces stronger evasive action does not necessarily result in a lower overall cost. Emergency braking, for example, may reduce the risk to a pedestrian while increasing the probability of a rear-end collision. Similarly, a steering maneuver may reduce direct pedestrian risk but transfer risk to vehicles or road users in an adjacent area.

The results therefore support the view that ethical conflict in autonomous driving is primarily a problem of constrained risk distribution. The objective is not simply to minimize one form of risk, but to compare how different feasible actions redistribute residual risk

among multiple participants. This interpretation is also consistent with the ethical-conflict analysis presented in Section 1, where occupant protection, public safety, rule compliance, and harm minimization were identified as potentially competing considerations. The variation in the optimal maneuver across the tested conditions further suggests that an ethical decision rule should remain sensitive to the surrounding traffic state rather than prescribe a fixed response for all emergency situations.

4.2. Responsibility Contributions and Risk Allocation

The responsibility model's results provide a second perspective on the same conflict. Under the baseline condition, the external traffic participant received the highest relative responsibility weight because the pedestrian's red-light crossing was the main trigger of the hazardous situation. The autonomous-driving system nevertheless retained a substantial responsibility contribution because it remained responsible for risk recognition, decision planning, and control execution. This distinction is important because the actor that initiates a hazardous situation is not necessarily the only actor that influences how the event develops.

The proposed model separates these contributions by considering perception and recognition, decision planning, control execution, human–machine interactions, traffic rules, environmental constraints, and data traceability. In this structure, a traffic-rule violation and a subsequent system response represent different parts of the event process. A pedestrian may contribute strongly to the initiation of risk, while the automated-driving system may still contribute to the eventual outcome through the adequacy of perception, planning, and control. The responsibility weights therefore represent relative contributions to the development of the event rather than mutually exclusive assignments of fault.

This distinction also has implications for maneuver selection. A decision that protects one party by transferring substantial risk to another party with little or no contribution to the original conflict may be difficult to justify solely on the basis of collision avoidance. The responsibility-transfer component of the ethical-cost function provides a mechanism for incorporating this consideration into the decision process. As a result, the framework does not treat responsibility only as a post-event quantity; it also considers whether a candidate maneuver redistributes risk in a manner that is inconsistent with the relative contributions of the involved parties.

At the same time, the calculated responsibility weights should not be interpreted as legal findings. The model provides engineering indicators of relative contribution based on scenario variables and predefined responsibility factors. Legal liability depends on additional legal standards, evidence, jurisdictional rules, and institutional considerations that are outside the scope of the present model. This distinction is consistent with the study's original formulation, which treats moral responsibility, causal contribution, and legal liability as related but non-equivalent concepts.

4.3. Linking Prospective Decision-Making with Responsibility Assessment

The framework can take vehicle-state estimates and predicted road-user trajectories from perception and prediction modules, evaluate feasible maneuvers, and send the selected maneuver to trajectory planning and control. The decision inputs, cost components, and responsibility weights can be logged for post-event review.

The prospective component evaluates feasible candidate maneuvers using the comprehensive ethical-cost function. The retrospective component maps responsibility-related factors to the autonomous-driving system, driver or safety operator, external traffic participants, road-environment or management actors, and manufacturer, developer, or operating entity. These two components are connected through responsibility transfer, which en-

ters the ethical decision cost together with collision risk, harm to life, rule violation, and vehicle stability.

This connection improves the traceability of the decision process. A selected maneuver can be examined not only in terms of its total cost but also in terms of the risk components and responsibility-related factors that contributed to that cost. Conversely, the responsibility assessment can be related back to the system states, traffic-rule conditions, human–machine interaction, and data records that were available during the event. This is particularly relevant for automated-driving systems because crash formation may involve multiple technical and human processes rather than a single driver action.

The framework should therefore be understood as an engineering method for maintaining consistency between decision evaluation and post-event analysis. It does not eliminate normative judgment, because the weighting of ethical costs and responsibility factors still reflects explicit assumptions about the relative importance of safety, rule compliance, stability, and risk transfer. However, these assumptions are represented quantitatively and can be reviewed, adjusted, and compared across scenarios. Li et al. [37] integrated Bayesian, equality, and maximin principles into reinforcement learning for highway merging, using data to infer context-dependent ethical weights. This learning-based approach complements the explicit modeling of multi-actor responsibility contributions in the present framework.

4.4. Limitations and Future Work

Several limitations should be considered when interpreting the present results. First, the validation was conducted using a representative pedestrian red-light crossing scenario. The five operating conditions share the same conflict structure. An application to other scenarios would require adapting the inputs and candidate maneuvers and recalibrating the responsibility-factor and ethical-cost weights. Future work will evaluate right-turn vehicle–pedestrian conflicts at signalized intersections, considering turning geometry and signal phasing.

Second, the responsibility-factor weights and ethical-cost weights are currently determined using expert elicitation and the analytic hierarchy process. This provides a structured method for assigning weights, but the resulting values remain dependent on expert judgment and scenario assumptions. The manuscript already recognizes that responsibility-factor weights should be adjusted according to the crash trigger, system operating state, traffic-participant behavior, and road conditions. Future work will extend the sensitivity analysis to responsibility-factor weights and additional operating conditions and refine weight calibration using empirical crash data, naturalistic driving data, or larger expert panels.

Third, the present evaluation is based on SCANeR–MATLAB co-simulation. The simulation environment enables controlled and repeatable comparison of candidate maneuvers, but it does not represent all uncertainties encountered in real traffic. Future work will propagate uncertainties in perceived positions and velocities, together with tracking delays, into TTC and collision-risk calculations and evaluate their effects on maneuver selection through simulation.

Finally, the current method evaluates a predefined set of candidate maneuvers, including emergency braking, leftward avoidance, rightward avoidance, and deceleration while maintaining the lane. It therefore addresses maneuver selection rather than continuous trajectory generation. Within model predictive controls, the five cost terms could be evaluated over the prediction horizon to optimize continuous steering and braking inputs subject to vehicle-dynamics constraints. Such an extension would also allow the framework to evaluate a broader range of feasible responses under more complex traffic interactions.

5. Conclusions

This study developed a responsibility-aware decision-making framework for ethical conflicts in autonomous driving by integrating a multi-actor responsibility model with a comprehensive ethical-cost function. The framework links the evaluation of candidate maneuvers with the assessment of relative responsibility contributions using scenario variables related to vehicle state, the road environment, traffic participants, human–machine interaction, and data traceability. SCANeR–MATLAB co-simulation was used to evaluate the framework in a pedestrian red-light crossing scenario. Under the baseline condition, emergency braking achieved the lowest comprehensive ethical cost of 0.276, while the external traffic participant and autonomous-driving system received relative responsibility weights of 0.50 and 0.26, respectively. Across the five operating conditions, the preferred maneuver varied between emergency braking and deceleration while maintaining the lane. Under Condition 4, with a following-vehicle gap of 10 m, lane-keeping deceleration achieved a lower comprehensive ethical cost than emergency braking (0.392 versus 0.410), illustrating how rear-end collision risks can influence maneuver selection.

Author Contributions: Conceptualization, H.S. (Hao Shi), C.S. and W.L.; methodology, H.S. (Hao Shi), C.S. and S.X.; software, H.S. (Hao Shi) and H.L.; validation, C.S., S.X., H.L., Y.C. and H.S. (Haiming Sun); formal analysis, H.S. (Hao Shi), Y.C. and H.S. (Haiming Sun); investigation, H.L.; resources, C.S. and W.L.; data curation, H.L.; visualization, H.S. (Hao Shi); writing—original draft preparation, H.S. (Hao Shi); writing—review and editing, C.S., W.L., S.X., Y.C. and H.S. (Haiming Sun); supervision, C.S., W.L., Y.Y. and S.X.; project administration, W.L. and Y.Y.; funding acquisition, C.S. and S.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China—grant number 2023YFB4302600; the Science and Technology Program of Suzhou—grant number SYG2025116; the Hubei Science and Technology Talent Service Enterprise Project—grant number 2026DJB020; the Hubei Science and Technology Project—grant numbers 2025CDB015 and 2024BAB087; and the Huanggang Technical Breakthrough General Project—grant number GGYB20260002.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to restrictions related to the experimental platform and data management.

Conflicts of Interest: Authors Dr. Haoran Li and Dr. Yang Yu were employed by the Hubei Tuopu Baide Technology Development Co., Ltd., Huanggang, China, and the author Dr. Haiming Sun was employed by Hubei Zhongcheng Industrial Technology Research Institute Co., Ltd., Shiyan 442000, China. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HMI	Human–Machine Interaction
TTC	Time to Collision
UAV	Unmanned Aerial Vehicle

References

1. Badue, C.; Guidolini, R.; Carneiro, R.V.; Azevedo, P.; Cardoso, V.B.; Forechi, A.; Jesus, L.; Berriel, R.; Paixão, T.M.; Mutz, F.; et al. Self-driving cars: A survey. *Expert Syst. Appl.* **2021**, *165*, 113816. [CrossRef]
2. Kiran, B.R.; Sobh, I.; Talpaert, V.; Mannion, P.; Sallab, A.A.A.; Yogamani, S.; Perez, P. Deep Reinforcement Learning for Autonomous Driving: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 4909–4926. [CrossRef]
3. Schwarting, W.; Alonso-Mora, J.; Rus, D. Planning and decision-making for autonomous vehicles. *Annu. Rev. Control Robot. Auton. Syst.* **2018**, *1*, 187–210. [CrossRef]

4. Koopman, P.; Wagner, M. Autonomous Vehicle Safety: An Interdisciplinary Challenge. *IEEE Intell. Transp. Syst. Mag.* **2017**, *9*, 90–96. [CrossRef]
5. Macioszek, E.; Tumminello, M.L. Simulating vehicle-to-vehicle communication at roundabouts. *Transp. Probl.* **2024**, *19*, 45–57. [CrossRef]
6. Di, X.; Dawid, H.; Muehlheusser, G. When engineering meets economics: AI-powered safe and accountable autonomous driving. *Artif. Intell. Transp.* **2025**, *2*, 100020. [CrossRef]
7. Sanchi, M. Systemic explainability for autonomous driving: Connecting ethics to law and society. *AI Soc.* **2026**, *41*, 10275–10293. [CrossRef]
8. Bonnefon, J.-F.; Shariff, A.; Rahwan, I. The social dilemma of autonomous vehicles. *Science* **2016**, *352*, 1573–1576. [CrossRef]
9. Awad, E.; Dsouza, S.; Kim, R.; Schulz, J.; Henrich, J.; Shariff, A.; Bonnefon, J.-F.; Rahwan, I. The Moral Machine experiment. *Nature* **2018**, *563*, 59–64. [CrossRef]
10. Himmelreich, J. Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations. *Ethical Theory Moral Pract.* **2018**, *21*, 669–684. [CrossRef]
11. Krügel, S.; Uhl, M. The risk ethics of autonomous vehicles: An empirical approach. *Sci. Rep.* **2024**, *14*, 960. [CrossRef]
12. Feess, E.; Muehlheusser, G. Autonomous Vehicles: Moral dilemmas and adoption incentives. *Transp. Res. Part B Methodol.* **2024**, *181*, 102894. [CrossRef]
13. Cecchini, D.; Dubljević, V. Moral Complexity in Traffic: Advancing the ADC Model for Automated Driving Systems. *Sci. Eng. Ethics* **2025**, *31*, 5. [CrossRef]
14. Goodall, N.J. Ethical Decision Making during Automated Vehicle Crashes. *Transp. Res. Rec.* **2014**, *2424*, 58–65. [CrossRef]
15. Geisslinger, M.; Poszler, F.; Lienkamp, M. An ethical trajectory planning algorithm for autonomous vehicles. *Nat. Mach. Intell.* **2023**, *5*, 137–144. [CrossRef]
16. Poszler, F.; Geisslinger, M.; Betz, J.; Lütge, C. Applying ethical theories to the decision-making of self-driving vehicles: A systematic review and integration of the literature. *Technol. Soc.* **2023**, *75*, 102350. [CrossRef]
17. Kirchmair, L.; Paulo, N. Taking ethics seriously in AV trajectory planning algorithms. *Nat. Mach. Intell.* **2023**, *5*, 814–815. [CrossRef]
18. Poszler, F.; Geisslinger, M.; Lütge, C. Ethical Decision-Making for Self-Driving Vehicles: A Proposed Model & List of Value-Laden Terms that Warrant (Technical) Specification. *Sci. Eng. Ethics* **2024**, *30*, 47. [CrossRef]
19. Gros, C.; Kester, L.; Martens, M.; Werkhoven, P. Addressing ethical challenges in automated vehicles: Bridging the gap with hybrid AI and augmented utilitarianism. *AI Ethics* **2025**, *5*, 2757–2770. [CrossRef]
20. Gros, C.; Werkhoven, P.; Kester, L.; Martens, M. A methodology for ethical decision-making in automated vehicles. *AI Soc.* **2025**, *40*, 6245–6256. [CrossRef]
21. Gros, C.; Martens, M.; Aliman, N.-M.; Kester, L.; Werkhoven, P. Ethical frameworks for automated vehicles: A systematic analysis and design. *AI Ethics* **2025**, *5*, 6041–6062. [CrossRef]
22. Vakili, E.; Amirkhani, A.; Mashadi, B. DQN-based ethical decision-making for self-driving cars in unavoidable crashes: An applied ethical knob. *Expert Syst. Appl.* **2024**, *255*, 124569. [CrossRef]
23. Yang, Y.; Wang, Q.; Wang, H.; Shen, J.; Ma, C.; Nie, B. An ethical decision-making framework for autonomous vehicles based on public moral preferences. *Accid. Anal. Prev.* **2025**, *222*, 108241. [CrossRef]
24. Milford, S.R.; Simone Elger, B.; Shaw, D. Bearing the weight: A qualitative study on expert views on integrating ethics in autonomous vehicles. *Transp. Res. Interdiscip. Perspect.* **2024**, *25*, 101096. [CrossRef]
25. Awad, E.; Levine, S.; Kleiman-Weiner, M.; Dsouza, S.; Tenenbaum, J.B.; Shariff, A.; Bonnefon, J.-F.; Rahwan, I. Drivers are blamed more than their automated cars when both make mistakes. *Nat. Hum. Behav.* **2020**, *4*, 134–143. [CrossRef]
26. Beckers, N.; Siebert, L.C.; Bruijnes, M.; Jonker, C.; Abbink, D. Drivers of partially automated vehicles are blamed for crashes that they cannot reasonably avoid. *Sci. Rep.* **2022**, *12*, 16193. [CrossRef]
27. Zhai, S.; Wang, L.; Liu, P. Not in Control, but Liable? Attributing Human Responsibility for Fully Automated Vehicle Accidents. *Engineering* **2024**, *33*, 121–132. [CrossRef]
28. Mu, Y.; Karasawa, M. Blame attribution and intentionality perception of human versus robot drivers: Implications for judgments about autonomous vehicles in moral dilemma contexts. *Cogent Psychol.* **2024**, *11*, 2384298. [CrossRef]
29. Vallor, S.; Vierkant, T. Find the Gap: AI, Responsible Agency and Vulnerability. *Minds Mach.* **2024**, *34*, 20. [CrossRef]
30. Conradie, N.H.; Nagel, S.K. Taking responsibility for the outcomes of autonomous technologies. *Ethics Inf. Technol.* **2025**, *27*, 21. [CrossRef]
31. Kirchmair, L. How to Regulate Moral Dilemmas Involving Self-Driving Cars: The 2021 German Act on Autonomous Driving, the Trolley Problem, and the Search for a Role Model. *Ger. Law J.* **2023**, *24*, 1184–1208. [CrossRef]
32. Shahini, F.; Zahabi, M. Effects of levels of automation and non-driving related tasks on driver performance and workload: A review of literature and meta-analysis. *Appl. Ergon.* **2022**, *104*, 103824. [CrossRef]

33. Hu, W.; Zhang, T.; Zhang, Y.; Chan, A.H.S. Non-driving-related tasks and drivers' takeover time: A meta-analysis. *Transp. Res. Part F Traffic Psychol. Behav.* **2024**, *103*, 623–637. [CrossRef]
34. Jin, L.; Liu, X.; Guo, B.; Han, Z.; Wang, Y.; Cao, Y.; Yang, X.; Shi, J. Impact of non-driving related task types, request modalities, and automation on driver takeover: A meta-analysis. *Saf. Sci.* **2025**, *181*, 106704. [CrossRef]
35. Chen, H.; Zhao, X.; Chen, C.; Li, Z.; Li, H.; Wang, Q. A systematic review on test performance of the driver takeover process in automated driving. *Accid. Anal. Prev.* **2025**, *215*, 108012. [CrossRef]
36. Nasser, G.; Morrison, B.W.; Wiggins, M.W. Mental models and driver takeover in automated vehicles: A systematic review. *Ergonomics* **2026**, *69*, 1292–1305. [CrossRef]
37. Li, L.; Nan, T.; Gan, J.; Zhang, D.; Qu, X.; Ran, B. Ethical Decision-Making in Autonomous Vehicles: A Reinforcement Learning Approach for Fair Risk Management. *Transp. Res. Part C Emerg. Technol.* **2026**, *190*, 105761. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.