vehicles

MDPI

Article

A Hybrid Decision-Making Framework for Autonomous Vehicles in Urban Environments Based on Multi-Agent Reinforcement Learning with Explainable AI

Ameni Ellouze [1], Mohamed Karray [2,*] and Mohamed Ksantini [3]

[1] Control and Energy Management Laboratory CEM Lab, ENIS, University of Sfax, Sfax 3018, Tunisia; ameni.ellouze@ipeis.usf.tn
[2] ESME Research Laboratory, Ivry Sur Seine, 94200 Paris, France
[3] Advanced Technologies for Image and Signal Processing (ATISP) Lab, ENETCom, University of Sfax, Sfax 3018, Tunisia; mohamed.ksantini@ipeis.usf.tn
* Correspondence: mohamed.karray@esme.fr

Abstract

Autonomous vehicles (AVs) are expected to operate safely and efficiently in complex urban environments characterized by dynamic and uncertain elements such as pedestrians, cyclists and adverse weather. Although current neural network-based decision-making algorithms, fuzzy logic and reinforcement learning have shown promise, they often struggle to handle ambiguous situations, such as partially hidden road signs or unpredictable human behavior. This paper proposes a new hybrid decision-making framework combining multi-agent reinforcement learning (MARL) and explainable artificial intelligence (XAI) to improve robustness, adaptability and transparency. Each agent of the MARL architecture is specialized in a specific sub-task (e.g., obstacle avoidance, trajectory planning, intention prediction), enabling modular and cooperative learning. XAI techniques are integrated to provide interpretable rationales for decisions, facilitating human understanding and regulatory compliance. The proposed system will be validated using CARLA simulator, combined with reference data, to demonstrate improved performance in safety-critical and ambiguous driving scenarios.

Keywords: autonomous vehicles; complex urban environment; hybrid decision; MARL; XAI; CARLA simulator

Check for updates

Academic Editors: Xubin Sun and Weifeng Zhong

Received: 31 October 2025
Revised: 12 December 2025
Accepted: 30 December 2025
Published: 2 January 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

Road traffic is constantly increasing, leading to a rise in accidents despite improvements in infrastructure [1]. Autonomous vehicles have emerged as a promising solution for improving safety and reducing traffic congestion. Urban areas, with their high traffic density, diversity of users and complex infrastructure, serve as ideal testing ground [2]. These vehicles, capable of moving without human intervention thanks to sensors, algorithms and artificial intelligence, represent a major advance in the field of transport [3]. Their potential in terms of safety, efficiency and accessibility is significant; however, their deployment remains hampered by several challenges, notably the reliability of their perception systems.

Despite the significant progress achieved in recent years in the field of autonomous mobility, the largescale implementation of these systems remains hampered by the complexity of modern urban environments. Autonomous vehicles, despite their potential to

improve safety and traffic flow, raise difficulties related to uncoordinated decisions, ambiguous signals and lack of transparency. These systems can aggravate certain problems, particularly in the case of uncoordinated behaviours in dense environments, incorrect interpretation of ambiguous signals, or lack of transparency detected, which hinders their acceptability and audibility.

In these environments, autonomous vehicles must be able to make safe, quick and adaptable decisions despite the presence of sudden scenarios such as: inattentive pedestrians, cyclists suddenly changing direction, intersections obscured by vehicles, congested or poorly signposted intersections, or even adverse weather conditions [4–7]. These challenges highlight the need to enhance the perception and decision-making capabilities of AVs to ensure their safe and effective integration into urban traffic.

The most common decision-making algorithms currently rely on techniques such as deep neural networks and fuzzy logic as well as reinforcement learning [8–14]. Although effective in well-defined scenarios, these approaches reach their limits when faced with high levels of uncertainty or human behaviour that is difficult to anticipate. Moreover, the lack of transparency raises ethical and regulatory concerns, especially in situations where it is essential to understand, legitimize, and evaluate the decisions made by an autonomous system.

The proposed framework addresses these limitations by combining multi-agent reinforcement learning to encourage cooperative decision-making, explainability tools to make vehicle choices understandable, and uncertainty-management mechanisms to reduce risky behavior in poorly perceived situations. Reinforcement learning is an AI paradigm in which an agent learns through trial and error to interact with its environment [15,16]. The multi-agent context, meanwhile, better reflects urban reality for cooperative traffic scenarios as several vehicles learn in parallel [17]. The overall objective is not only to improve technical performance, but also to strengthen the reliability, acceptability and regulatory compliance of large-scale autonomous systems.

Therefore, the contribution of this work is to introduce an original hybrid decision-making framework that coherently combines MARL and XAI in order to simultaneously meet the performance, robustness and transparency requirements of autonomous driving systems. Consequently, the main objective is to develop a modular and cooperative decision-making system in which each agent specialises in a crucial sub-task, such as intention prediction, obstacle avoidance or trajectory planning. The decisions of these agents are integrated cooperatively in order to achieve an optimal compromise between safety, efficiency and interpretability.

The native integration of XAI modules, based in particular on the SHAP method, guarantees operational transparency by generating intelligible explanations of the decisions made by agents. For example, if an autonomous vehicle chooses to stop at a specific location, the XAI module allows it to quantify the relative contribution of the different variables explaining this choice, such as the proximity of a pedestrian or a red light. This ability to interpret decisions at the local and global levels promotes transparency, provides an essential auditing tool, and strengthens compliance with growing regulatory requirements for algorithmic accountability [18,19].

The proposed approach is distinguished by three major contributions: (1) a hierarchical arbitration mechanism that dynamically prioritises decisions between agents in the event of conflict; (2) native integration of explainability into the decision-making process, avoiding weak post hoc analyses; and (3) a multi-objective reward function that balances safety, efficiency and transparency. To validate this framework, experiments were conducted in the CARLA (Car Learning to Act) simulator [20,21] on critical urban scenarios. Performance, evaluated using metrics such as collision rate and decision latency, as well as

an explainability indicator based on the SHAP score, shows significant gains in terms of collision reduction, responsiveness and stability of explanations. The proposed framework thus represents an integrated advance in the design of autonomous systems that are reliable, interpretable and compliant with real-world constraints.

The remaining of this paper is structured as follows: Section 2 reviews the related works. Section 3 summarizes our methodology based on the proposed approach combined the MARL and the XAI. Section 4 is devoted to the experimental results and discussion. The last section provides the main conclusions and suggests some future perspectives.

2. Related Works

Autonomous driving in urban environments poses a major challenge for intelligent decision-making systems. Historically, rule-based systems have been favoured for their simplicity and interpretability. However, their rigidity and limited capacity for generalisation constrain their effectiveness in complex and dynamic urban environments. To overcome these limitations, AI-based approaches particularly deep learning, imitation learning, and neuro-symbolic reasoning have emerged as more flexible and adaptive alternatives [8,9].

Deep learning methods have enabled significant advances thanks to convolutional and recurrent neural network architectures capable of end-to-end driving based on raw sensory data [18]. Models such as NVIDIA's PilotNet and CARLA integrations have demonstrated the feasibility of real-time trajectory prediction in simulated environments [7,22]. However, the opaque nature of these networks raises concerns about interpretability and safety, particularly in ambiguous urban contexts. Furthermore, their dependence on large annotated datasets makes them costly to deploy and sometimes difficult to generalise to unfamiliar conditions.

In [23], the authors presented a new decision-making framework, UrbanDrivePathway, which leverages deep learning, reinforcement learning, and real-time traffic analysis to improve VA performance. This framework integrates EfficientDet for object detection (97.69% accuracy), graph neural networks (GNNs) for trajectory prediction (97.42% accuracy), and Monte Carlo tree search (MCTS) for decision-making (97.38% accuracy). Model predictive control (MPC) is used for trajectory planning, with an accuracy of 97.61%, while an eco-optimised driving strategy improves energy efficiency with an accuracy of 97.69% in energy consumption analysis. The study envisages future improvements that will focus on multi-agent collaboration, adversarial scenario training and advanced explainability techniques.

The authors in [24] proposed TransFuser, a model based on a single-agent architecture with multi-sensor fusion using transformers. This approach allows for the effective integration of visual and depth information to improve the perception of autonomous vehicles.

The end-to-end approach proposed in [25] is based on the implicit affordance's method, a reinforcement learning approach applied to urban driving. This technique enables an agent to manage the trajectory, avoid pedestrians and vehicles, and detect traffic lights. This is the first RL agent capable of handling such complex scenarios, including traffic light management. The model's effectiveness was confirmed by its victory in the Camera Only track of the CARLA Challenge.

In the recent state of the art on controlling autonomous vehicles connected via MARL, the majority of contributions focus on inter-vehicle collaboration, distributed planning or coordination at intersections. Where related work focuses on performance (e.g., minimizing travel time or improving throughput at intersections). In [26], the authors define a decentralized MARL problem for navigating through such intersections and introduces a new advanced algorithm, Multi-Agent Game-Aware Deep Deterministic Policy Gradient (MA-GA-DDPG), based on the DDPG with attention mechanism and level-k, allowing to improve safety and efficiency at complex intersections, by finely modeling multi-agent

interactions. However, this work remains focused on behavioral optimization, without providing a transparent mechanism to interpret the decisions of agents.

The authors in [27] proposed MARL to improve coordination between autonomous vehicles in shared environments. Given that these systems are vulnerable to adversarial perturbations that may lead to critical errors, the authors propose R-CCMARL, a robust algorithm integrating a universal policy, risk estimation, and constraint formulation aimed at maximising performance even in worst-case scenarios. Simulation experiments demonstrate the system's robustness against attacks while preserving high operational efficiency.

In [28], the authors model lane change decision making for multiple autonomous vehicles (coexisting with "human" vehicles) as a MARL problem. They use a MA2C architecture (actor critic multi agent), a parameter sharing, and a multi objective reward function (security, comfort, efficiency). They show that their approach improves efficiency and safety in multiple traffic densities.

The work presented in [29] focuses on robustness and safety in MARL applied to autonomous driving. They propose two algorithms (CSQ and CS MADDPG) based on a Stackelberg-type model and bi-level optimization, tested on realistic scenarios (junctions, roundabouts, intersections). They report improvements in terms of security and reward compared to classic MARL baselines.

However, these approaches are often limited to behavioural optimization functions, without explicit integration of explicability mechanisms. For example, in [30], the authors integrated reinforcement learning and XAI for traffic jam recognition and prediction. The complexity of urban traffic in static rule-based systems can be tackled by using a dynamic approach that combines RL for adaptive learning and XAI for transparent decision-making. The proposed method improves safety, privacy, and prediction accuracy, achieving an impressive accuracy rate of 98.10% while significantly reducing the rate of missing data.

Our contribution in this area is based on a new paradigm that integrates multi-agent reinforcement learning with explainable artificial intelligence approaches. This hybrid decision-making framework aims to optimise the coordination and responsiveness of autonomous vehicles in highly complex urban environments. By integrating these approaches, decision-making is both more effective and robust, and transparency and trust are enhanced through the intelligent agent's ability to interpret and justify actions.

3. Methodology

This section describes the proposed methodology based on the MARL and the XAI frameworks.

3.1. MARL Framework

3.1.1. Description

In Multi-Agent Reinforcement Learning, several agents interact in the same environment and learn to make decisions simultaneously. This approach is particularly useful in autonomous driving scenarios, where MARL models can therefore be used to coordinate lane changes, negotiate intersections and drive by treating each vehicle as an agent with shared or individual rewards [17]. Unlike in the single-agent case, each agent influences the evolution of the environment and therefore the learning of others, creating a non-stationary framework.

In this work, the MARL model proposes three main agents each specialised in a specific task in order to reflect the distributed and collaborative nature of the systems embedded in AVs.

- Agent 1: Intention prediction: dedicated to anticipating the behaviour of pedestrians, cyclists and other vulnerable road users.

- Agent 2: Dynamic obstacle avoidance: responsible for detecting and managing interactions with moving obstacles.
- Agent 3: Optimal trajectory planning: determine the safest and most efficient trajectory while considering the environment and dynamic constraints.

Each agent receives input from sensors (cameras, LiDAR, radar) and makes an assessment (ambiguous areas, incomplete detection, unreliable sensor), which allows them to take risks and uncertainties into account in their decisions by providing interpretable explanations (e.g., why a trajectory was avoided or an action prioritised). Cooperative decisions are then merged and commands are sent to the vehicle's actuators. Finally, a human supervision module allows real-time consultation of decisions and XAI explanations. Figure 1 describes the architecture of the hybrid decision for autonomous driving.

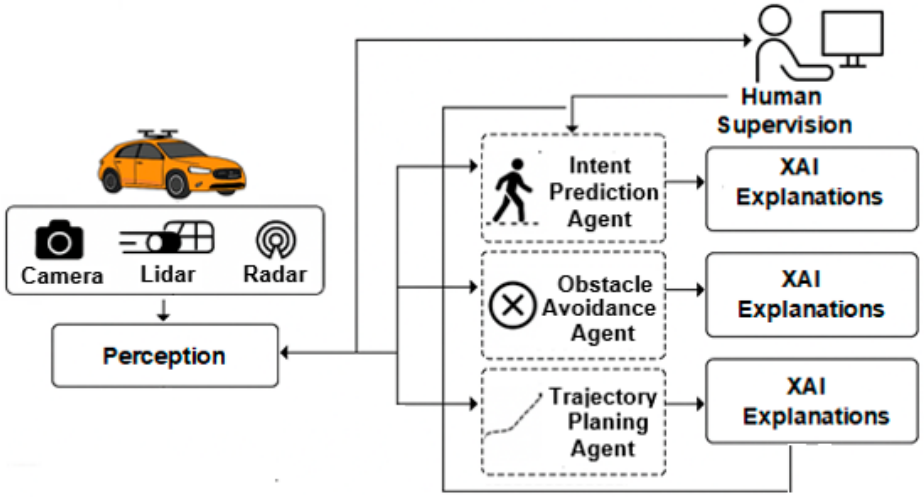


Figure 1. Hybrid Decision architecture for Autonomous Driving in Complex Urban Environments.

3.1.2. MARL Formulation Problem

The MARL problem is formulated as a modelled multi-agent stochastic game, which can be summarised in the following steps [31]:

- N represents the set of agents $i \in \{1, \ldots, N\}$
- Each agent i chooses an action a_i according to a policy $\Pi_i(a_i|o_i)$, where a_i corresponds to the action space of agent i and o_i is the partial observation of the global state.
- The environment evolves according to the transition function $P(s\prime|s, a_1, \ldots, a_N)$ where s is the space of global states (position, speed, traffic signals, predicted intentions).
- Every agent receives the reward function r_i corresponding to its objectives (avoid a collision, minimise travel time, save fuel, etc.).
- The goal is to maximise the cumulative expected reward over all time steps t, taking into account the evolution of the system over time:

$$\mathbb{E}\left[\sum_t \gamma^t r_i^t\right] \quad (1)$$

where $\gamma \in [0, 1]$ is the discount factor controlling the importance of future rewards.

This formalism adopts a Centralised Training, Decentralised Execution (CTDE) approach, which allows agents to train collectively while making local decisions during the inference phase [32]. During the centralised training phase, agents $(a_1, \ldots, a_N)$ have centralised access to all global information S (states, actions and rewards), enabling them to model collective dynamics more accurately and learn to coordinate their behaviour. This allows the construction of a global critic that estimates the value of a joint action set

in a given state: $Q_i(s|, a_1, \ldots, a_N)$. Thus, the critic provides a more stable update of the individual policies Π_i as it incorporates interactions between agents.

However, During the execution phase, the critics are omitted, and each agent acts solely on the basis of its partial observations (each agent retains only its local policy $\Pi_i(a_i|o_i)$) without relying on central supervision or information outside its scope. This mechanism ensures that agents can operate autonomously in a real environment, while having learned beforehand to collectively optimise their policies.

Potential conflicts between the specialised agents are managed using a hierarchical arbitration mechanism guided by the priority of objectives and the level of uncertainty estimated by each agent. Concretely, the obstacle avoidance agent takes priority over the others, as it is responsible for immediate safety. Then comes the planning agent, which optimises efficiency (speed, comfort) when the situation is considered safe. Finally, the intention prediction agent provides anticipatory recommendations but does not immediately block the action; it influences trajectory or avoidance adjustments. This hierarchy can be explained as follows:

- Each agent calculates an uncertainty score σ_i via policy entropy compared to a threshold θ_i set empirically after simulation tests.
- If the first avoidance agent detects a risk ($\sigma_1 > \theta_1$), its action is imposed.
- Otherwise ($\sigma_2 < \theta_2$), the second planning agent takes the action.
- If neither agent is prioritised, the prediction agent influences an action.
- The final policy is then assigned to the agent with the highest priority.

3.2. XAI Framework

XAI plays an essential role in improving the decision-making of perception systems, allowing both technical experts and non-specialists to understand how decisions are made, how errors are detected and what is the meaning of the results produced by AI [33]. In this context of rapid evolution of autonomous vehicle technologies, AI transparency is essential to ensure security, trust, and robustness. In our hybrid decision-making system, where several specialised MARL agents interact in a dynamic urban environment, it is essential not only to make effective and safe decisions, but also to be able to explain each action taken. In this context, the integration of an XAI module is essential for transparency and traceability. Among the existing XAI approaches, we have chosen SHAP because of its theoretical rigour, explanatory granularity, and compatibility with the deep architectures used in our MARL framework. Unlike methods such as LIME, which rely on local approximations that are often sensitive to noise, SHAP is based on cooperative game theory and Shapley values. This guarantees two fundamental properties: additivity (the sum of the individual contributions of the variables is equal to the model output) and consistency (a variable whose influence increases sees its SHAP value increase). Indeed, after each action, SHAP identifies the input variables that contributed to this decision such as the distance to a pedestrian, the state of the traffic light or the perceived level of uncertainty [19]. For example, when an agent decides to brake, the XAI module may reveal that this action is mainly due to a red light and the proximity of a pedestrian. This ability to generate local explanations makes it possible to finely interpret the system's behaviour. Unlike classical MARL approaches, often perceived as "black boxes", our architecture offers decision-making transparency that strengthens both the robustness, human trust and traceability of algorithmic choices. The use of SHAP in our framework is therefore not an arbitrary choice, but responds to a methodological requirement for rigorous and robust explainability, natively integrated into the decision-making architecture given by Figure 2.

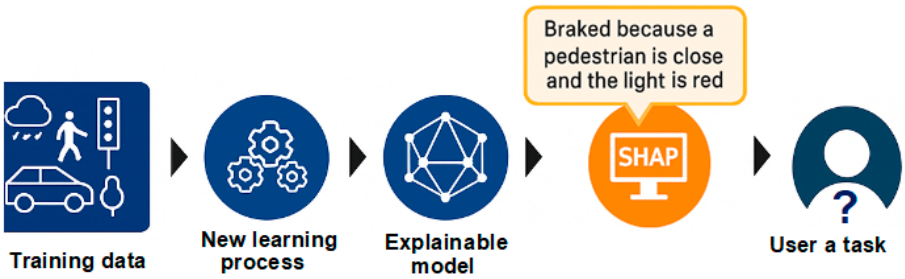


Figure 2. XAI process.

SHAP values are calculated at each decision step, enabling traceability and interpretation of the decisions made by each agent. More specifically, the value functions $Q_i(s|, a_1, \ldots, a_N)$ learned by the agents are considered predictive models whose outputs must be explained in terms of the variables of the observed state. The principle can be reformulated as follows:

Consider an observed state $x = (x_1, \ldots, x_M)$ composed of M characteristics (e.g., inter-vehicle distances, presence of pedestrians, speed limits, traffic light status). The contribution of each characteristic x_j to an agent's action decision is defined by a Shapley value [19]:

$$\varnothing_j = \sum_{S \subseteq \{1,\ldots,M\} \setminus \{j\}} \frac{|S|!(M - |S| - 1)!}{M!} \left[f_{S \cup \{j\}} \left(x_{S \cup \{j\}} \right) - f_S(x_S) \right] \quad (2)$$

where S represents a subset of characteristics, and $f_S(x_S)$ corresponds to the model's prediction when only the variables in S are taken into account. In our case and in relation to MARL, the predictive function f is assimilated to the agent's value-action function:

$$Q_i(\mathrm{s}, a_i) \approx \varnothing_0 + \sum_{j=1}^{M} \varnothing_j \quad (3)$$

where $\varnothing_0$ corresponds to the baseline value and $\varnothing_j$ to the marginal contribution of characteristic x_j. This additive decomposition makes it possible to understand, for each action decision, which environmental factors had the greatest impact.

4. Results

This section is dedicated to presenting the results from the simulations in terms of metrics and performance.

4.1. Experimental Environment

The training and the evaluation of the MARL model are performed within interactive simulation environment where the agent can test its decisions and receive feedback through rewards. Unlike static datasets, in this project, the experiments were carried out in CARLA simulator using the Python API (python version 3.7, CARLA-PyAPI), establishing an environment well-suited for training and benchmarking AI models [20,21].

Figure 3 provides a clear description of the implementation phase, detailing the operation of the proposed system and its three main modules:

1. CARLA-based simulation environment.
2. Decision-making module based on MARL agents.
3. XAI module based on SHAP.

The open-source urban driving CARLA is used to handle complex urban configurations, various traffic agents and dynamic weather conditions. The chosen environment provides realistic city maps, simulated on-board sensors (LIDAR, camera, GPS) and configurable scenarios for controlled experimentation. This simulator records trajectories,

states (position, speed), decisions (brakes, steering wheel, accelerator), and events (traffic lights, roads, pedestrians) in structured files (JSON, CSV, etc.). The three previously described agents are tested simultaneously to simulate realistic traffic behaviour. Four critical scenarios are defined for this purpose:

1. Unexpected pedestrian crossing
2. Vehicle blocked in the middle of an intersection
3. Missing or incorrect traffic signs
4. Changing weather conditions (rain, fog)

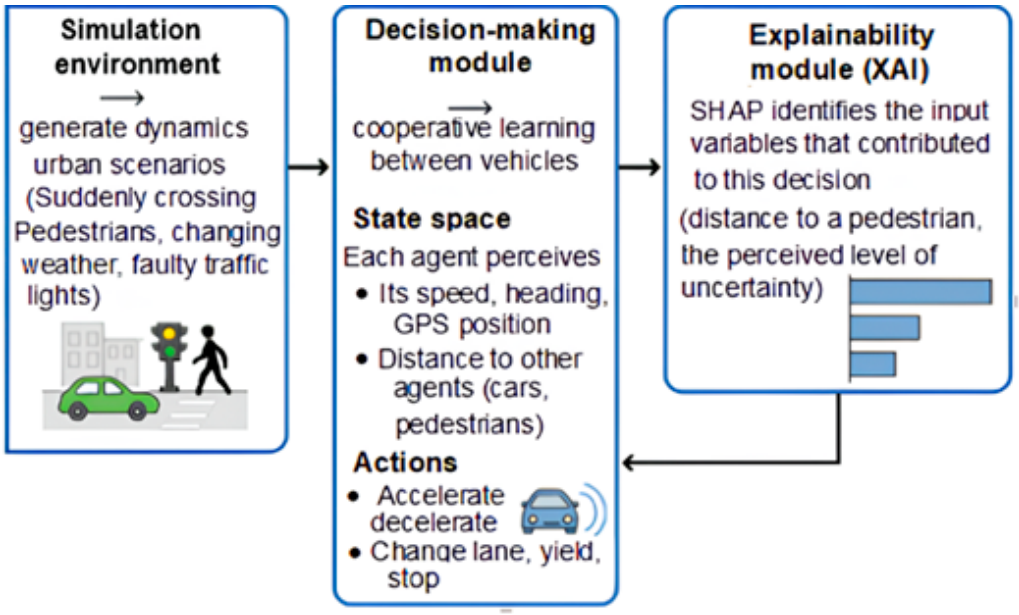


Figure 3. Description of the implementation phase.

Each scenario is repeated 50 times with different initial conditions (random seeds, distinct placements of pedestrians/vehicles). A total of 200 episodes were run. An episode is considered successful if the vehicle reaches its target destination without collision or extended obstruction lasting more than 10 s.

The following Table 1 lists all the features available in the sensors integrated into CARLA simulator.

Features of the used sensors:

Table 1. Features of the sensors integrated into CARLA simulator.

Sensor	Parameters	Role
Front RGB camera	Resolution = 800 × 600 px, FOV = 90°, frequency = 20 Hz, position = (x = 1.5 m, z = 2.4 m)	Used for detecting pedestrians and vehicles. Scenarios involving unexpected pedestrians and missing signage rely heavily on this camera. Detection errors account for some collisions.
Rear RGB camera	Resolution = 800 × 600 px, FOV = 110°, frequency = 20 Hz	Detects vehicles approaching from behind at intersections. Indirect influence on safety (less critical than frontal safety).
3D LiDAR	32 channels, range = 50 m, frequency = 10 Hz, Gaussian noise σ = 0.05 m	Contributes to obstacle detection and trajectory prediction. Essential in scenarios involving blocked vehicles and changing weather conditions. Reduced performance in fog/rain explains the increase in collision rates.

Table 1. *Cont.*

Sensor	Parameters	Role
Front radar	Range = 20 m, angular aperture = 20°, frequency = 20 Hz	Detects the relative speed of moving objects. Helps reduce collisions.
GNSS	Simulated accuracy = ±0.5 m, frequency = 1 Hz	Used for global localisation and trajectory planning. Not directly related to collisions but influences the stability of path following.
IMU	Accelerometer noise = 0.05 m/s^2, gyroscope = 0.005 rad/s, frequency = 100 Hz	Provides vehicle dynamics. Influences the calculation of decision-making latency and driving comfort.
Collision sensor	Continuous activation	Sensor used to measure the collision rate.
Lane invasion sensor	Continuous activation	Detects line crossings. Used to calculate the success rate (%).

The criteria and metrics performances are presented in the following section.

4.2. Evaluation Criteria and Metrics

Performance is evaluated according to three metrics:
Safety:

- Collision rate (%): the average frequency of collisions per scenario.
- Collisions/100 km-AV: the number of collisions relative to the total distance travelled by the three AVs.
- Success rate (%): the proportion of episodes without major incidents.

Efficiency:

- Decision Latency (ms): the average time required for the agent to generate an action. This is linked to the complexity of policy Π_i, which increases when the perceived states (camera + noisy lidar) raise processing time.

Transparency:

- SHAP score (0–1): the accuracy and stability of the explanations generated $\varnothing_j$ to justify decisions. Good decision readability is achieved under normal conditions, but decreases in ambiguous scenarios.

In order to ensure the effectiveness of each MARL agent in dealing with complex driving and environmental conditions, R1, R2 and R3 rewards are defined for each agent in relation to the evaluation metrics:

- R1: The obstacle avoidance agent receives a reward that heavily penalises collisions and low-time situations. This formulation is directly related to the collision rate measured experimentally.
- R2: The path planning agent is rewarded for reaching the destination quickly, measured by the distance travelled along the target, and penalised for unnecessary stops or excessive detours. The reward also includes a penalty proportional to the difference between the instantaneous speed and the target speed. These terms are directly related to the decision latency and success rate metrics.
- R3: The intent prediction agent is rewarded for correctly anticipating movements, assessed using measures such as the average displacement error (ADE) and final displacement error (FDE). At the same time, explainability techniques (SHAP) are applied retrospectively to assess the stability and interpretability of decisions.

The global reward function R is therefore formulated as a weighted combination that prioritises safety in the evaluation of system performance. However, the integration of efficiency and transparency avoids an imbalance that could compromise overall performance.

In other words, even though safety dominates the reward function, the contributions of efficiency and transparency ensure balanced optimisation, promoting reliability, operational speed and decision traceability. The function R is given as follows:

$$R = 0.5 \times R_1 + 0.3 \times R_2 + 0.2 \times R_3 \quad (4)$$

For a more detailed explanation, the weights assigned to the three components of the reward function safety (0.5), efficiency (0.3) and explainability (0.2) are the result of a compromise between normative requirements, functional constraints and empirical evaluation. Safety, weighted at 0.5, is prioritised in accordance with safety standards (ISO 26262, SAE J3016) [34] which require strict risk minimisation in driving environments. Efficiency, with a weighting of 0.3, ensures smooth and realistic behaviour, which is essential for continuous urban traffic. Transparency, introduced through XAI mechanisms via SHAP score, is weighted at 0.2 to enhance the interpretability of decisions. Following tests carried out on simulated scenarios, it was found that too high a weighting on efficiency leads to aggressive behaviour, while too low a weighting on explainability alters the consistency of explanations.

It should be noted that all performance metrics in this section are standardised as they can naturally have values in very different ranges.

4.3. Results and Discussion

Before presenting the quantitative results of the metrics, we introduce several experimental illustrations from simulations carried out with our CARLA environment. These figures illustrate scenes before and after the intervention of the MARL system.

Figure 4 illustrates a foot passenger begins to cross the road at an unmarked crossing as the vehicle approaches, with perfect visibility but a critical situation due to proximity. After making a decision, the vehicle stops or slows down significantly before the crossing line, demonstrating the activation of the obstacle avoidance agent, priority according to hierarchical arbitration.



(**a**) (**b**)

Figure 4. Scene of unexpected pedestrian crossing (green cercle). (**a**) Scene before the intervention of the MARL; (**b**) Scene after the decision of the MARL.

Figure 5 illustrates that the autonomous vehicle operates in an urban area under dense fog. We note the absence of clearly visible vertical or illuminated signs at the intersection. The vehicle continued on its path with caution. This decision reflects proactive management of uncertainty through the intention prediction module. The decision to move forward slowly and steadily shows that the MARL system was able to handle this ambiguity.



(**a**) (**b**)

Figure 5. Scene of missing traffic signs and degraded perception. (**a**) Scene before the intervention of the MARL; (**b**) Scene after the decision of the MARL agents.

Figure 6 illustrates the scenario depicting an urban intersection in dense fog conditions, with no visible signage or vehicles on the ground, as seen from an onboard camera. After making its decision, the system maintains its trajectory without modification, indicating that the camera position has not changed significantly. This means that the system has determined that it is safe to stay on course without changing direction or stopping abruptly.



(**a**) (**b**)

Figure 6. Scene of urban intersection with onboard camera. (**a**) Scene before the intervention of the MARL; (**b**) Scene after the decision of the MARL.

Table 2 presents the results of the various criteria and metrics evaluated in the different scenarios.

Table 2. Results of the performance criteria and metrics.

Scenario	Collisions/100 km-AV	Decision Latency (ms)	SHAP Score	Success Rate (%)
Unexpected Pedestrian	2.8	120	0.82	90
Blocked Vehicle	3.6	128	0.79	85
Missing/Faulty Signal	5.1	135	0.75	78
Changing Weather	7.0	145	0.71	80
Average	4.6	132	0.77	83

Based on Table 2, the results of the various measurements are interpreted in the following points:

Safety: The collision rate remains relatively low in the unexpected pedestrian scenario (2.8 collisions/100 km-AV), confirming the ability of the intention prediction agent to handle sudden human behaviour. Performance decreases when the context is unclear, such as in the absence of, or with incorrect signage (5.1 collisions per 100 km-AV) or under changing weather conditions (7.0 collisions per 100 km-AV). The trend suggests that even though skilled agents are capable of managing isolated subtasks, the system's overall safety is still impacted by perception degradation and ambiguity in traffic rules.

Efficiency: Decision latency remains below 150 ms in all scenarios (average = 132 ms), which is generally considered safe in urban environments. The slight increase in latency in adverse weather conditions (145 ms) is explained by the increased load related to sensor noise and reduced visibility. It is important to note that latency remains stable between agents, demonstrating that the modular MARL architecture does not introduce computational bottlenecks, even in dense traffic.

Transparency: SHAP scores, ranging from 0.71 to 0.82, indicate that the explanations produced are both coherent and easily interpretable.

Transparency is higher in well-defined scenarios (unexpected pedestrian), while it decreases in cases of regulatory or perceptual ambiguities (absent signage, poor weather). Nevertheless, scores above 0.7 indicate that the integration of XAI techniques already provides understandable justifications for human users, facilitating regulatory compliance and building trust.

In summary, we can conclude that the most favourable scenario corresponds to the appearance of an unexpected pedestrian, which results in few collisions, reduced latency, a high SHAP score and a success rate of 90%. In contrast, the most unfavourable scenario occurs when weather conditions change, resulting in an increase in collisions, maximum latency and a low SHAP score. The two intermediate scenarios also confirm that the absence of reliable information (missing signage) or the presence of an obstacle (immobilised vehicle) significantly degrades the system's performance.

Thus, the hierarchical arbitration mechanism has demonstrated its relevance for road observation (obstacle at 15 m, density 0.7), the intention agent which displays the best overall performances (SHAP score 0.71–0.82, success rate 78–90%) presents a low uncertainty ($\sigma = 0.35 < \theta = 0.5$) and is therefore validated as a reliable source. Conversely, the safety officer, whose performance drops in critical scenarios (7.0 collisions/100 km), has his proposal rejected here ($\sigma = 0.42 > \theta = 0.3$), which directly reflects his vulnerability to detection uncertainties. As for the trajectory agent, its high uncertainty ($\sigma = 0.50 > \theta = 0.4$) in this situation explains its lower success rates (78–85%) in complex environments.

Table 3 presents the reward values that show the contribution of each agent.

Table 3. Reward values.

Scenario	R_1	R_2	R_3	Global R
Unexpected Pedestrian	0.88	1.00	1.00	0.94
Blocked Vehicle	0.80	0.80	0.85	0.82
Missing/Faulty Signal	0.65	0.60	0.60	0.64
Changing Weather	0.50	0.00	0.00	0.35
Average	0.71	0.60	0.61	0.69

The analysis of standardised rewards highlights a clear hierarchy among the different scenarios tested. The unexpected pedestrian crossing scenario achieves the best global performance (R = 0.94), reflecting an optimal combination of safety (R1 = 0.88), decision-

making efficiency (R2 = 1) and transparency (R3 = 1). In contrast, changing weather conditions are the most unfavourable case, with a reduced overall reward (R = 0.35) due to the simultaneous deterioration of safety, decision-making speed and clarity of explanations provided. Intermediate scenarios, such as a blocked vehicle (R = 0.82) and absent or erroneous signage (R = 0.64), illustrate a gradual decline in performance linked respectively to increased coordination complexity and environmental changes. The obtained results validate the effectiveness of the proposed reward function in capturing the three critical dimensions (safety, efficiency, transparency) and allow the robustness of the system to be assessed in a variety of contexts.

The following figures illustrate all the results discussed above.

Figure 7 illustrates the average number of collisions per 100 km travelled by an autonomous vehicle (AV).

- The unexpected pedestrian scenario presents the best result, with only 2.8 collisions/100 km-AV, demonstrating the responsiveness and robustness of the avoidance agent.
- The blocked vehicle remains under control (3.6), confirming the system's ability to handle static obstacles.
- On the other hand, the scenarios of erroneous signaling (5.1) and changing weather (7.0) show a strong increase in collisions, linked respectively to the ambiguity of the signaling and to the degradation of the sensors under rain/fog. This confirms that decreases R1 as soon as the environmental uncertainty increases.

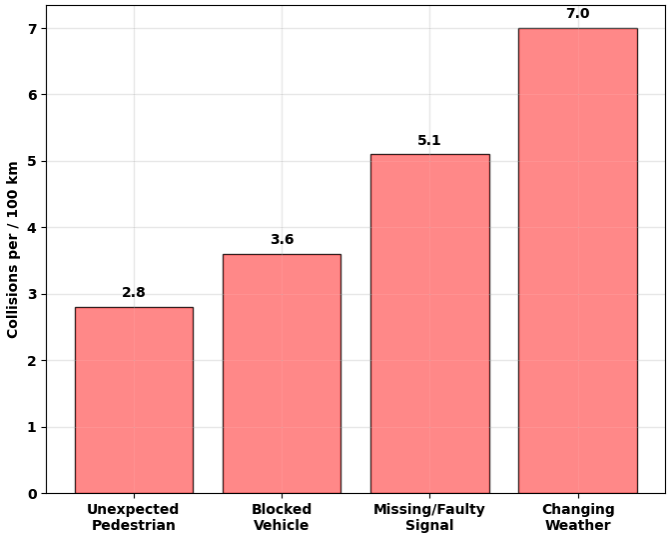


Figure 7. Collision rates by scenario.

Figure 8 illustrates the latency (ms) by measuring the average time required to produce an action (acceleration, braking, or steering).

- The lowest values are observed for the unexpected pedestrian scenario (120 ms), a sign of a quick and effective decision.
- The latency increases slightly for the blocked vehicle (128 ms) and faulty signaling (135 ms).
- A maximum latency of 145 ms is observed under changing weather conditions, as a direct consequence of increased sensor noise. At the same time, the efficiency reward decreases from 1.0 to 0.0 as the complexity of the environment increases.

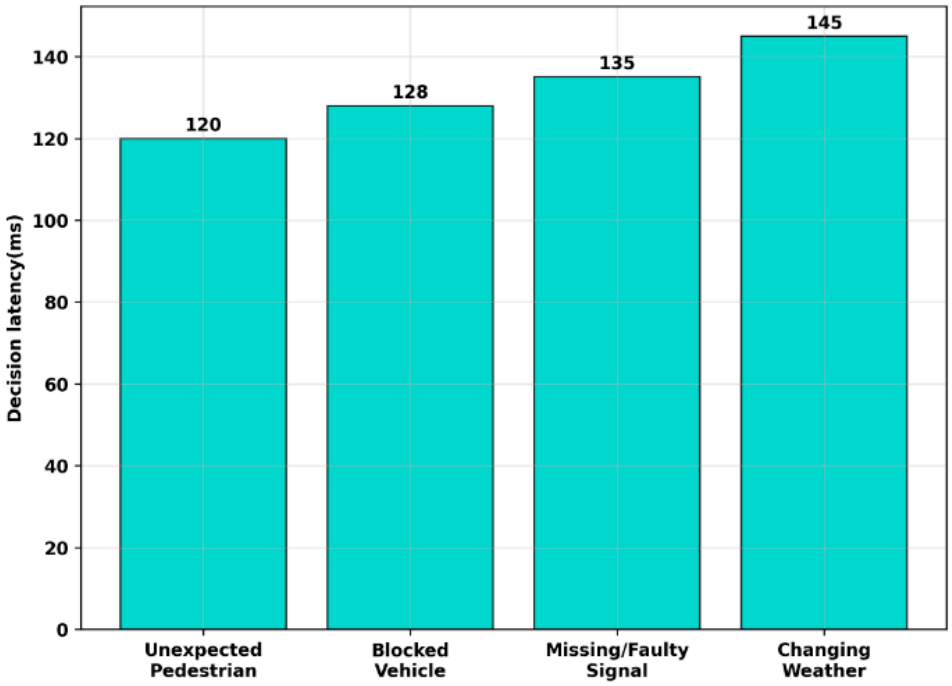


Figure 8. Decision latency by scenario.

Figure 9 represents the transparency and stability of decisions through the SHAP score (0–1).

- The single scenarios of unexpected pedestrian and blocked vehicle have high scores (0.82 and 0.79), indicating that decisions are based on logical characteristics (frontal obstacle, distance, relative speed).
- Ambiguous scenarios involving erroneous signalling and changing weather have lower scores (0.75 and 0.71), reflecting a less clear interpretation in which several sensors contribute to the decision in a scattered manner. These variations confirm that the transparency R deteriorates when perception becomes uncertain, which affects the traceability and trust of the system.

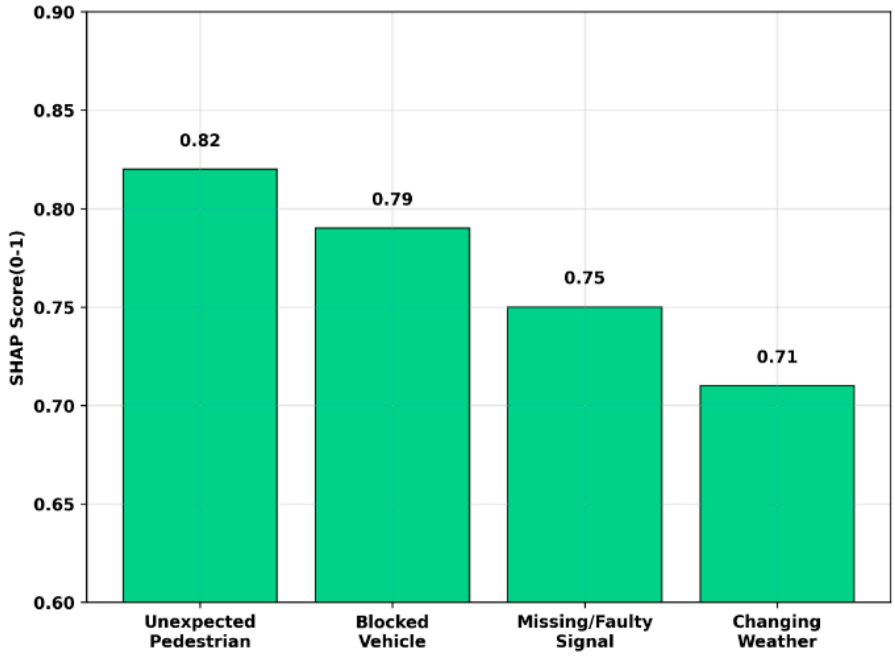


Figure 9. Average SHAP score by scenario.

Figure 10 illustrates that the success rate is:

- Highest for unexpected pedestrian (90%) and blocked vehicle (85%) demonstrating good inter-agent coordination and stable learning.
- Reduced to 78% for erroneous signalling and 80% for changing weather, where conditions disrupt decision-making. This figure clearly illustrates the direct relationship between sensory stability and overall performance, validated by the evolution of cumulative reward.

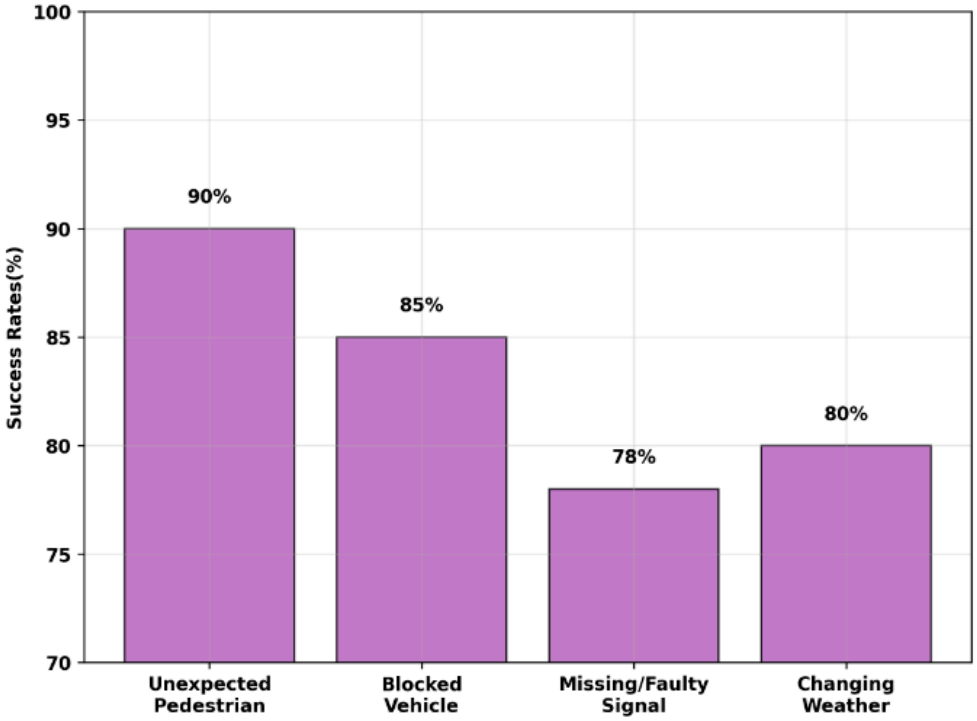


Figure 10. Overall success rate.

The evaluation of the robustness of our MARL-XAI architecture was initially carried out in the four critical scenarios covering various urban situations. These first results showed a good compromise between security, decision latency and explainability, with an average success rate higher than 83% and a SHAP score reflecting a stable explicability above 0.75 for most cases. However, we recognize that the evaluation of scalability remains an essential area for improvement, but in order to consolidate these initial results, we tried to extend our work and validate it by adding three additional scenarios, targeting more complex and critical urban situations: a complex intersection with multiple intersecting priorities, a case of cooperative merger approaching a bottleneck, and the unforeseen intervention of an emergency vehicle. The results of the different metrics are presented in Table 4.

Table 4. Performance criteria and metrics results of the added scenarios.

Scenario	Collisions/ 100 km-AV	Decision Latency (ms)	SHAP Score	Success Rate (%)
Intersection convergence	5.8	150	0.68	78
Cooperative merging	4.5	142	0.72	81
Emergency braking	3.2	138	0.76	87

The new scenarios show degraded performance in complex situations (lower SHAP, higher latency) but maintain good safety, reflecting the real challenges of autonomous driving in a dense environment. Indeed, the complex intersection scenario revealed a

high decision-making challenge with a collision rate of 5.8 per 100 km and a latency of 155 ms, reflecting increased demand on arbitration modules and a relative decrease in the SHAP score (0.68) indicating less stable explainability. On the other hand, in the case of cooperative fusion, the performances are more balanced, demonstrating the effectiveness of the intent prediction module in multi-agent interactions. Finally, the scenario involving an emergency vehicle confirms the system's ability to quickly prioritize safety with a high SHAP score (0.76), evidence of clear and justified decision-making.

In general, the analysis of the seven scenarios shows that, despite the complexification of decisions to be made and the intensification of interactions between agents, our system manages to maintain a satisfactory level of performance. These results confirm the ability of our architecture to adapt to dynamic environments, which aims to extend this work to the multidimensional agent and multiple scenarios.

4.4. Comparison of Results

Compared with existing studies, this work provides substantial added value by combining MARL and XAI within a unified framework, while targeting complex and ambiguous urban scenarios. This combination not only strengthens the robustness and the explicability of the system, but also enables a rigorous evaluation based on precise quantitative metrics (collision rate, decision latency and SHAP score).

Although many studies address these techniques separately or qualitatively manner, this approach limits the possibility of fair comparison, since a few provide results that are quantitatively comparable to the performance achieved in this work.

The model developed in [24] considers a single-agent multi-sensor fusion architecture based on transformers. It achieves an average success rate of 90% on the CARLA Leaderboard with a latency of 150 to 170 ms and a collision rate of about 10 to 15/10 km. Experiments, also show that TransFuser achieves peak driving performance while reducing collisions by 76% compared with classical geometric fusion. But this method remains limited in its ability to explicitly manage uncertainty and to provide, however, a clear interpretation of decisions.

The end-to-end approach in [25], based on direct reinforcement learning with 'implicit affordances', achieves success rates between 75% and 85%, but has a higher collision rate (≈6–10/10 km) and latency close to 150 ms. Although this architecture offers valuable insights for advancing unsupervised learning, it remains vulnerable.

Furthermore, the work presented in [19] proposes a hybrid XAI system for autonomous vehicles combining LIME and SHAP in order to improve decision transparency. The author reports an inference latency of 280 ms for perception models (ResNet 18), which is a useful reference for embedded architectures. By comparison, our proposed system obtains an average decision latency of 132 ms, while integrating not only explicability but also uncertainty estimation and inter-agent coordination within a unified architecture.

Regarding the robustness against perturbations, the authors in [26] focus on a MARL framework designed to resist adversarial attacks within a cooperative driving environment. While their study demonstrates an improvement in safety at intersections simulated with CARLA, it does not provide precise metrics on collision rate, latency or the explainability of decisions. In contrast, our work introduces an explicit measure of interpretability via the SHAP score (average: 0.77) for each scenario. This contribution fills a gap often identified in systematic reviews, where most RL approaches remain "black boxes".

5. Conclusions

Given that autonomous driving in urban environments requires decision-making capabilities that are robust, explainable and adaptable to uncertainty, this work proposes

an innovative hybrid decision-making architecture combining multi-agent reinforcement learning (MARL) and explainable artificial intelligence (XAI), specifically designed for complex and ambiguous urban environments. Unlike traditional single-agent approaches, our system relies on three specialised agents (intention prediction, obstacle avoidance, trajectory planning) cooperating within a modular framework. The integration of XAI modules (SHAP) into each agent allows for the generation of interpretable local explanations, thereby enhancing transparency, traceability and compliance with regulatory requirements. One of the major contributions of this research is the implementation of a SHAP score in a multi-agent framework, which remains rare in the literature.

The evaluation simulation in ambiguous urban scenarios (unexpected pedestrians, faulty signage, variable weather conditions) demonstrates the robustness of the system with an average success rate of 83%, a lower latency (132 ms) and a reduced collision rate (4.6/100 km-AV), while introducing an average SHAP score of 0.77. This offers a quantitative measure of explainability absent from previous models. Our hybrid MARL and XAI framework combines the decision-making robustness of multi-agent reinforcement with measurable explainability, which gives it an advantage both in terms of safety (fewer collisions) and the traceability of decisions. This combination paves the way for more reliable autonomous driving systems that comply with regulatory transparency requirements.

The obtained results show a clear advantage in comparison with several recent works. Our analysis shows that while previous studies explore individual aspects of the problem such as MARL, robustness, or explainability, few have succeeded in integrating them all into a unified, quantitative and modular architecture, as proposed in our framework.

However, in the system showing high robustness in structured environments (pedestrians and clear obstacles), performance drops as soon as changes increase such as rain and absence of signals. This reveals some limitations of the framework in terms of proactive uncertainty management and coordination in conflict situations. Furthermore, the experiments were conducted with a limited number of agents (three) and scenarios (four), without validating scalability to larger fleets or coordination in congested intersections.

These limitations lead us to propose some perspectives which will consolidate the validity of the proposed MARL framework and make it a more generalisable solution:

- Explicitly integrating uncertainty estimation and management modules into the MARL architecture.
- Extending the system to a larger number of agents and scenarios to test its scalability in denser and more interactive environments.
- Expand the evaluations to include multiple algorithms of MARL in order to establish a robust benchmark.

Author Contributions: All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. World Health Organization (WHO). *Road Traffic Injuries*; World Health Organization: Geneva, Switzerland, 2024.
2. Soica, A.; Gheorghe, C. Revolutionizing Urban Mobility: A Systematic Review of AI, IoT, and Predictive Analytics in Adaptive Traffic Control Systems. *Electronics* **2025**, *14*, 719. [CrossRef]

3. Gondhalekar, G.; Mathiyalagan, P. Artificial Intelligence Applications in Autonomous Vehicles: Navigating the Future of Transportation Systems. In Proceedings of the ITM Web of Conferences, Enugu, Nigeria, 4–5 December 2025; Volume 76. [CrossRef]
4. Rajesh, B.; Ramakrishna, D.; Raju, A.R.; Chavan, A. Object detection and classification for autonomous vehicle. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2021; Volume 1817. [CrossRef]
5. Turki, A.; Kahouli, O.; Saleh, A.; Ksantini, M.; Aloui, A.; Ben Amara, M. A sophisticated Drowsiness Detection System via Deep Transfer Learning for real time scenarios. *AIMS Math.* **2024**, *9*, 3211–3234. [CrossRef]
6. Feng, D.; Haase-Schütz, C.; Rosenbaum, L.; Hertlein, H.; Glaeser, C.; Timm, F.; Wiesbeck, W.; Dietmayer, K. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 1341–1360. [CrossRef]
7. Karray, M.; Triki, N.; Ksantin, M. A New Speed Limit Recognition Methodology Based on Ensemble Learning: Hardware Validation. *Comput. Mater. Contin.* **2024**, *80*, 119–138. [CrossRef]
8. Pentela, V.K.; Deepalakshmi, P. Improved Threat Intelligence and Real-Time Protection with Next-Generation Firewall Solutions. In Proceedings of the 2025 Global Conference in Emerging Technology (GINOTECH), Pune, India, 9–11 May 2025. [CrossRef]
9. Khan, H.; Chaudhari, T.; Ramesh, J.; Kranthi, A.S.; Muniyandy, E.; El-Ebiary, Y.; Devadhas, D. Neuro-Symbolic Reinforcement Learning for Context-Aware Decision Making in Safe Autonomous Vehicles. *Int. J. Adv. Res. Comput. Sci.* **2025**, *16*, 599–609. [CrossRef]
10. Ibrahum, A.D.M.; Hussain, M.; Hong, J.E. Deep Learning Adversarial Attacks and Defenses in Autonomous Vehicles: A Systematic Review from a Safety Perspective. *Artif. Intell. Rev.* **2024**, *58*, 28. [CrossRef]
11. Wu, X.; Chen, W.; Chen, P.C.Y.; Zhao, Z.; Liu, J. LSTM Network: A deep learning approach for short-term traffic forecast. *IET Intell. Transp. Syst.* **2017**, *11*, 68–75. [CrossRef]
12. Moridpour, S.; Akhtar, M. A review of traffic congestion prediction using Artificial Intelligence. *J. Adv. Transp.* **2021**, *2021*, 8878011. [CrossRef]
13. Siswipraptini, P.C.; Aziza, R.N.; Hartanti, D. Optimization of smart traffic lights to prevent traffic congestion using fuzzy logic. *Telkomnika* **2019**, *17*, 320. [CrossRef]
14. Ranjan, N.; Bhandari, S.; Zhao, H.P.; Kim, H.; Khan, P. City-Wide Traffic congestion prediction based on CNN, LSTM, and Transpose CNN. *IEEE Access* **2020**, *8*, 81606–81620. [CrossRef]
15. Taghavifar, H.; Wei, G.; Taghavifar, L. Socially Intelligent Reinforcement Learning for Optimal Automated Vehicle Control in Traffic Scenarios. *IEEE Trans. Autom. Sci. Eng.* **2024**, *22*, 129–140. [CrossRef]
16. Ben Elallid, B.; Bagaa, B.; Benamar, M.; Mrani, N. A reinforcement learning based autonomous vehicle control in diverse daytime and weather scenarios. *J. Intell. Transp. Syst.* **2024**, *29*, 1–14. [CrossRef]
17. Hua, M.; Qi, X.; Chen, D.; Jiang, K.; Liu, Z.E.; Sun, H.; Zhou, Q.; Xu, H. Multi-Agent Reinforcement Learning for Connected and Automated Vehicles Control: Recent Advancements and Future Prospects. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 16266–16286. [CrossRef]
18. Alotaibi, J. Enhancing Traffic Accident Severity Prediction: Feature Identification Using Explainable AI. *Vehicles* **2025**, *7*, 38. [CrossRef]
19. Tahir, H.A.; Alayed, W.; Hassan, W.; Haider, A. A Novel Hybrid XAI Solution for Autonomous Vehicles: Real-Time Interpretability Through LIME–SHAP Integration. *Sensors* **2024**, *24*, 676. [CrossRef]
20. Dosovitskiy, A.; Ros, G.; Codevilla, F.; Lopez, A.; Koltun, V. CARLA: An open urban driving simulator. In Proceedings of the 1st Annual Conference on Robot Learning, Mountain View, CA, USA, 13–15 November 2017; Volume 78, pp. 1–16.
21. Malik, S.; Khan, M.A.; El-Sayed, H. CARLA: Car Learning to Act—An Inside Out. *Procedia Comput. Sci.* **2022**, *198*, 742–749. [CrossRef]
22. Csato, A.; Mariasiu, F.; Csiki, G. Analysis of the Influence of Image Resolution in Traffic Lane Detection Using the CARLA Simulation Environment. *Vehicles* **2025**, *7*, 60. [CrossRef]
23. Ranga, J.; Prasath, A.; Kumar, N.; Naveenkumar, R.; Vadar, P.S.; Fiaz, S. UrbanDrivePathway: A Decision-Making Framework for Navigating Urban Autonomous Vehicles in Complex Traffic Systems. In Proceedings of the 2025 8th International Conference on Trends in Electronics and Informatics (ICOEI), Irunelveli, India, 24–26 April 2025. [CrossRef]
24. Prakash, A.; Chitta, K.; Geiger, A. Multi-Modal Fusion Transformer for End-to-End Autonomous Driving. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 7077–7087. [CrossRef]
25. Toromanoff, M.; Wirbel, E.; Moutarde, F. End-to-End Model-Free Reinforcement Learning for Urban Driving Using Implicit Affordances. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 7153–7162. [CrossRef]
26. Liu, J.; Hang, P.; Na, X.; Huang, C.; Sun, J. Cooperative decision-making for CAVs at unsignalized intersections: A Marl approach with attention and hierarchical game priors. *IEEE Trans. Intell. Transport. Syst.* **2022**, *26*, 443–456. [CrossRef]

27. Wang, C.; Wang, Z.; Aouf, N. Robust Multi-Agent Reinforcement Learning Against Adversarial Attacks for Cooperative Self-Driving Vehicles. *ET Radar Sonar Navig.* **2025**, *19*, e70033. [CrossRef]
28. Zhou, W.; Chen, D.; Yan, J.; Li, Z.; Yin, H.; Ge, W. Multi-Agent Reinforcement Learning for Cooperative Lane Changing of Connected and Autonomous Vehicles in Mixed Traffic. *Auton. Intell. Syst.* **2022**, *2*, 5. [CrossRef]
29. Zheng, Z.; Gu, S. Safe Multi-Agent Reinforcement Learning with Bilevel Optimization in Autonomous Driving Autonomous Vehicles in Mixed Traffic. *IEEE Trans. Artif. Intell.* **2025**, *6*, 829–842. [CrossRef]
30. Khan, S.; Ghazal, T.M.; Alyas, T.; Waqas, M.; Raza, M.A.; Ali, Q.; Khan, M.A.; Abbas, S. Towards Transparent Traffic Solutions: Reinforcement Learning and Explainable AI for Traffic Congestion. *Int. J. Adv. Comput. Sci. Appl.* **2025**, *16*, 503–511. [CrossRef]
31. Canese, L.; Cardarilli, G.; Nunzio, L.D.; Fazzolari, R.; Giardino, D.; Re, M.; Spanò, S. Multi-Agent Reinforcement Learning: A Review of Challenges and Applications. *Appl. Sci.* **2021**, *11*, 4948. [CrossRef]
32. Zhou, Y.; Liu, S.; Qing, Y.; Zheng, T.; Chen, K.; Song, J.; Song, M. CADP: Towards Better Centralized Learning for Decentralized Execution in MARL. In Proceedings of the 24th International Conference IJCAI, Detroit, MI, USA, 19–23 May 2025; pp. 7218–7226. [CrossRef]
33. Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G.Z. XAI—Explainable artificial intelligence. *Sci. Robot.* **2019**, *4*, eaay7120. [CrossRef] [PubMed]
34. *ISO26262-2:2011*; Road Vehicles–Functional Safety–Part2: Management of Functional Safety. International Organization of Standardization, ISO: Geneva, Switzerland, 2011.

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.