

RESPOND: Risk-Enhanced Structured Pattern for LLM-driven Online Node-level Decision-making in Autonomous Driving

Dan Chen¹, Heye Huang^{2,✉}, Tiantian Chen², Zheng Li³, Yongji Li⁴, Yuhui Xu⁵, Sikai Chen³

Cite this article: <https://doi.org/10.26599/COMMTR.2026.9640037>

ABSTRACT: Current LLM-based driving agents relying on unstructured plain-text memory suffer from low-precision scene retrieval and inefficient reflection. To address this, we present RESPOND, a structured decision-making framework for LLM agents grounded in risk patterns. RESPOND constructs a unified 5×3 ego-centric matrix that encodes spatial topology and road constraints, enabling consistent retrieval of spatial-risk configurations. Building on this, a hybrid Rule+LLM pipeline employs a two-tier memory lookup: exact patterns ensure rapid, safe action reuse in high-risk contexts, while sub-patterns facilitate personalized style adaptation under low risk. Furthermore, a pattern-aware reflection mechanism abstracts tactical corrections from crash frames to update structured memory, achieving "one-crash-to-generalize" learning. Experimental results demonstrate the effectiveness of this design. In highway-env, RESPOND surpasses state-of-the-art LLM-based and RL-based agents while generating substantially fewer collisions. With step-wise human feedback, the agent acquires a *Sporty* driving style within approximately 20 decision steps through sub-pattern abstraction. For real-world validation, we evaluate RESPOND on 53 high-risk cut-in scenarios extracted from the HighD dataset. For each event, we intervene at the moment immediately preceding the cut-in and allow RESPOND to re-decide the driving action. Compared to the recorded human behavior, RESPOND's decisions reduce subsequent risk in 84.9% of these scenarios, demonstrating the feasibility and practical relevance of RESPOND under real-world driving conditions. These results highlight the potential for real-world autonomous driving, personalized driving assistance, and proactive hazard mitigation. Code will be released at: <https://github.com/gisgrid/RESPOND>.

KEYWORDS: large language models (LLM); risk pattern representation; structured memory; reflection learning; autonomous driving

1 Introduction

Achieving safe and reliable autonomous driving in complex mixed-traffic environments requires decision-making systems that can retrieve past experiences precisely, adapt to new situations, and learn efficiently from sparse safety-critical events (Huang et al. 2025; Wang et al. 2021; Hao et al. 2024). Recent work has explored Large Language Models (LLMs) for driving decision-making, leveraging their reasoning and generalization capabilities to interpret diverse traffic situations (Mao et al. 2023; Zhou et al. 2026; Wen, Yang, et al. 2023). Despite these advances, current LLM-driven agents primarily store experiences as plain-text narratives, which limits their utility in several fundamental ways. First, scene retrieval becomes imprecise because semantic matching alone cannot reliably distinguish small but safety-critical variations in high-risk or long-tail events. Second, reflection learning is inefficient since crash cases stored in unstructured text make it difficult to extract reusable knowledge, restricting generalization from failures. Moreover, most existing approaches lack explicit encoding of spatial-risk dependencies between the ego vehicle and surrounding entities, making it difficult to represent the risk-perception patterns that human drivers rely on when making safety-critical decisions.

To address these limitations, we propose RESPOND, a knowledge-driven decision framework grounded in the quantitative Driver Risk

Field (DRF) (Kolekar et al. 2020; Huang, Wang, et al. 2020). RESPOND encodes each ego-centric scene as a unified 5×3 risk matrix that captures neighboring vehicles and road-edge constraints, yielding a discrete and comparable spatial-risk pattern for reliable memory retrieval and improved recall fidelity over plain-text storage. Based on this representation, RESPOND introduces a hybrid Rule+LLM decision pipeline with a two-tier memory structure: exact pattern matches enable high-confidence action reuse in high-risk contexts, while sub-pattern matches summarize essential spatial-risk configurations to support personalization under low risk (Kou et al. 2025; Zeng et al. 2025). A lightweight LLM refines decisions only when necessary, preserving interpretability and adaptability. In addition, a pattern-aware reflection mechanism enables one-crash-to-generalize learning by extracting corrected strategies from pre- and post-crash frames and writing them back as reusable sub-patterns, preventing repeated failures in unseen scenarios. The main contributions are summarized as follows:

1. We propose a unified risk-environment representation that encodes ego-centric traffic context into a DRF-based 5×3 risk matrix. This structured formulation yields a discrete and comparable pattern space that enables high-precision memory retrieval and improves over unstructured textual storage.
2. We design a two-tier decision framework that combines rule-based strategies with LLM reasoning. Exact pattern reuse

¹ School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China. ² Cho Chun Shik Graduate School of Mobility, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea. ³ Department of Civil and Environmental Engineering, University of Wisconsin–Madison, Madison WI 53706, USA. ⁴ Sun Yat-sen University, Guangzhou 510275, China. ⁵ China Mobile Group Guangdong Co., Ltd., Guangzhou 510623, China.

✉ Corresponding author. E-mail: heyehuang@kaist.ac.kr

Received: December 23, 2025; Revised: March 8, 2026; Accepted: June 22, 2026

© The Author(s) 2026. This is an open access article under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0, <http://creativecommons.org/licenses/by/4.0/>).

supports high-confidence actions in safety-critical scenarios, while sub-pattern abstraction enables flexible decision-making and efficient human-in-the-loop personalization under low risk.

3. We develop a pattern-aware reflection mechanism that analyzes pre- and post-crash frames to extract tactical corrections and writes them back into structured memory. This mechanism supports one-crash-to-generalize learning and accelerates knowledge accumulation when safety-critical data are limited.

The remainder of this paper is organized as follows. Section 2 reviews related works. Section 3 outlines the RESPOND framework. Section 4 details the methodology. Section 5 presents experiments and results on both simulation and real-world datasets. Section 6 concludes the paper and discusses limitations and future research directions.

2 Related Works

2.1. Quantitative Risk Modeling

Risk modeling is fundamental for ensuring safe and context-aware autonomous decision-making. Existing approaches can be broadly categorized into three classes.

(1) **Kinematics-based surrogate safety indicators.** Metrics such as Time-To-Collision (TTC) approximate short-term collision risk based on vehicle trajectory data (Munro et al. 2008; Peesapati et al. 2018; Cheng et al. 2025; Shalev-Shwartz et al. 2017). However, these methods are limited to pairwise interactions and fail to capture multi-agent spatial dependencies in complex traffic.

(2) **Probabilistic graphical risk models.** Approaches such as Bayesian networks and Partially Observable Markov Decision Processes (POMDPs) provide a principled framework for risk estimation under uncertainty (Lauri et al. 2022; Spaan 2012; Ross et al. 2011). However, they suffer from high computational complexity and limited scalability in real-time multi-agent environments.

(3) **Field-based risk modeling.** Classical Artificial Potential Field (APF) and Social Force models represent interactions via attractive and repulsive forces (Bounini et al. 2017; Helbing and Molnar 1995), but suffer from limited expressiveness for behavior reasoning. More recent approaches, such as the Driver Risk Field (DRF), model interaction risk as structured spatial fields by integrating kinematic and environmental factors (Kolekar et al. 2020; Wang et al. 2016; Huang, Wang, et al. 2020). These representations are well suited for structured encoding, as they naturally produce grid-like risk patterns. However, existing methods are typically stateless and lack mechanisms for storing and retrieving risk-action patterns for decision reuse.

2.2. LLM-Based Driving Agents

LLMs have recently gained traction in autonomous driving due to their ability to interpret complex scenarios and generate decision-oriented reasoning (Liu et al. 2023; Wang et al. 2025). Existing work can be broadly categorized into three directions.

(1) **Natural language-based scene understanding and high-level decision making.** Early studies use LLMs to process scene descriptions and generate high-level actions such as lane changes (Mao et al. 2023; Wen, Yang, et al. 2023). Multimodal extensions further incorporate visual inputs to enhance reasoning (Wang et al. 2023; Liu et al. 2025; Hou et al. 2025). However, these approaches rely on free-text representations, making it difficult to align and reuse decisions across structurally similar scenarios, especially in safety-

critical cases.

(2) **Memory-augmented and reflection-capable LLM agents.** Recent work introduces memory and self-reflection into LLM-driven decision-making. Systems such as DiLu and general reflection frameworks enable iterative improvement through feedback (Wen, Fu, et al. 2023; Zhou et al. 2026; Zhao et al. 2025). However, memory is typically stored in unstructured text, causing retrieval to depend on semantic similarity rather than structural correspondence, which limits precise recall in high-risk scenarios.

(3) **Personalized autonomous driving and human-in-the-loop adaptation.** Approaches such as PADriver and Drive-as-you-say enable style modulation through user preferences or demonstrations (Kou et al. 2025; Cui et al. 2024). While effective for behavior tuning, these methods lack structured representations of personalized strategies, making them difficult to generalize or reuse across different scenarios.

3 Framework

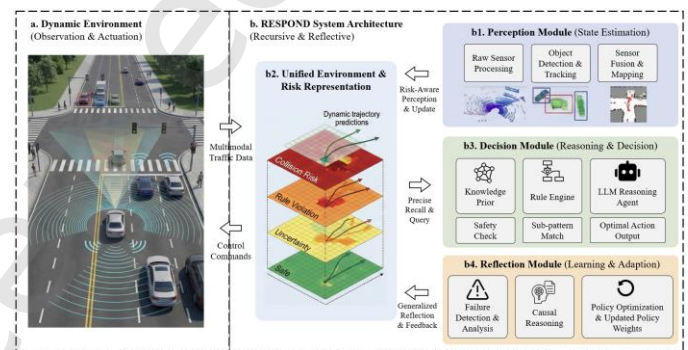


Fig. 1. Architecture of the RESPOND framework. The framework integrates dynamic environment observation, unified risk representation, hybrid decision-making, and reflection-based learning. The four core modules include perception, risk representation, decision-making, and reflection.

The RESPOND framework addresses the low retrieval fidelity and inefficient reflection of unstructured textual memory in LLM-based driving agents. As shown in Fig. 1, RESPOND converts ego-centric traffic observations into a structured risk representation through the Perception Module (b1) and Unified Environment & Risk Representation module (b2). Raw sensor or simulation states are mapped into a multi-layer representation covering collision risk, rule violation, and uncertainty, and are finally encoded as a unified 5×3 spatial-risk matrix. This matrix balances spatial coverage and compactness by extending a basic 3×3 neighborhood with two additional lateral rows, enabling near-adjacent lane interactions to be captured while maintaining efficient pattern retrieval.

Based on this representation, the Decision Module (b3) combines knowledge priors, rule-based constraints, sub-pattern matching, and LLM reasoning to generate safe actions. Familiar high-confidence patterns can be reused directly, while novel or complex cases are handled by the LLM and checked by safety rules. The Reflection Module (b4) further analyzes safety-critical events, identifies failure causes, and updates structured memory, enabling one-crash-to-generalize learning and reducing repeated unsafe behaviors.

4 Methodology

This section presents the methodology of RESPOND,

including unified risk–environment encoding, structured memory construction, hybrid decision-making, and pattern-aware reflection for safety improvement and personalization.

4.1. Unified Risk–Environment Encoding

RESPOND introduces a Unified Risk–Environment Encoding to convert heterogeneous driving information, including spatial topology, dynamic states, and quantitative risk, into a compact discrete representation for structured memory retrieval and LLM inference.

4.1.1 Hierarchical Scene Abstraction

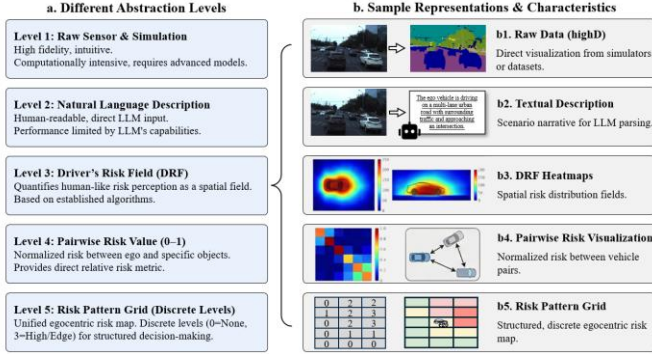


Fig. 2. Five abstraction levels for structured scene and risk representation in RESPOND, ranging from raw sensor or simulation data to textual descriptions, DRF heatmaps, pairwise risk values, and the final 5×3 discrete risk pattern grid.

Existing LLM-based driving agents face a mismatch between flexible language descriptions and quantitative safety reasoning. To address this, RESPOND adopts a five-level abstraction hierarchy (Fig. 2) that progressively transforms the same driving scene from raw sensor or simulation states into a structured decision unit. Level 1 captures high-fidelity vehicle states, Level 2 provides an LLM-readable textual description, Level 3 maps kinematic states into continuous DRF heatmaps, and Level 4 summarizes these fields into normalized pairwise risk values in $[0, 1]$. Finally, Level 5 constructs a unified 5×3 discrete risk pattern grid that encodes spatial topology and risk intensity for structured memory storage and decision-making.

4.1.2 Quantitative Risk Modeling

We adopt the DRF model as the quantitative basis for Level 4 and extend it to omnidirectional ego-centric interactions. For a surrounding vehicle j , the risk value RV_j is defined as the normalized overlap between the risk fields of the ego vehicle and vehicle j :

$$RV_j = \frac{A_{\text{overlap}}(\text{ego}, j)}{A_{\text{ref}}} \quad (1)$$

where A_{overlap} denotes the intersection area of the two risk fields projected onto the road plane. The continuous DRF intensity field $F(x, y)$ is converted into an effective geometric region using a fixed threshold $\theta = 0.01$, defined as $F(x, y) \geq \theta$. The overlap area $A_{\text{overlap}}(\text{ego}, j)$ is then computed as the intersection of the effective regions of the ego and surrounding vehicles.

This thresholding converts the continuous DRF field into a binary spatial region for consistent comparison across vehicles. The normalization term A_{ref} ensures that $0 \leq RV_j \leq 1$. In practice, moderate variations of θ (e.g., $\theta \in [0.005, 0.02]$) have minimal impact on the

resulting discrete risk patterns.

The normalization term A_{ref} is defined according to the interaction direction:

$$A_{\text{ref}} = \begin{cases} A_{\text{ego}}, & \text{for rear and rear-adjacent interactions} \\ A_j, & \text{for front and front-adjacent interactions} \end{cases} \quad (2)$$

where A_{ego} and A_j denote the risk field areas of the ego and surrounding vehicle, respectively. This direction-aware normalization captures asymmetric interactions, particularly in rear-following scenarios. Under this definition, $RV_j = 1$ when the effective risk region of the ego fully contains that of vehicle j , corresponding to high-risk interactions such as aggressive cut-in or rapid closing. To account for multi-lane interactions, the ego DRF is laterally shifted to compute risk values in six directional zones: front, rear, left-front, left-rear, right-front, and right-rear.

4.1.3 Discrete Matrix Representation

To transform continuous risk values into a retrieval-ready representation, we discretize the spatial context into a grid matrix $P \in \{0, 1, 2, 3\}^{5 \times 3}$. The matrix is centered on the ego vehicle, covering the ego lane and two adjacent lanes with five longitudinal segments. Each element $P_{i,j}$ represents the discretized risk level of a spatial cell based on RV :

$$P_{i,j} = \begin{cases} 0, & \text{if } RV_{i,j} < 0.34 \quad (\text{Safe}) \\ 1, & \text{if } 0.34 \leq RV_{i,j} < 0.66 \quad (\text{Attention}) \\ 2, & \text{if } 0.66 \leq RV_{i,j} < 0.99 \quad (\text{Danger}) \\ 3, & \text{if } RV_{i,j} = 1 \quad (\text{Critical}) \end{cases} \quad (3)$$

To capture safety constraints that may be underestimated by DRF, we introduce two deterministic overrides. First, road-edge cells are assigned $P_{i,j} = 1$ to encode non-drivable regions. Second, near-field risk in the center column is computed using proximity and TTC instead of DRF overlap.

For longitudinal interactions in the same lane, discretization follows a rule-based procedure. If the distance is ≥ 30 m, the risk is set to 0 (Safe). Otherwise, TTC is estimated based on relative motion. If $\text{TTC} \leq 1$ s or distance ≤ 15 m, the risk is set to 3 (Critical); otherwise, it is set to 1 (Attention). The same procedure is applied symmetrically to leading and following vehicles, and the final center-column value is obtained by combining front and rear risks, capped at 3. For adjacent lanes, the ego vehicle is laterally shifted and the same procedure is applied.

The resulting 5×3 risk pattern provides a compact and standardized representation that preserves critical safety information while enabling efficient memory retrieval and stable LLM reasoning.

The thresholds are chosen based on standard safety criteria and empirical observations in highway driving. Specifically, $\text{TTC} \leq 1$ s indicates imminent collision risk, while distance thresholds (15 m and 30 m) capture near-field interactions. The lateral masking threshold $\tau_{\text{lat}} = 0.67$ aligns with the boundary between Attention and Danger levels in Eq. (3).

All thresholds (30 m, 15 m, and $\text{TTC} \leq 1$ s) are fixed and consistently applied across all experiments.

4.2. Structured Memory Construction

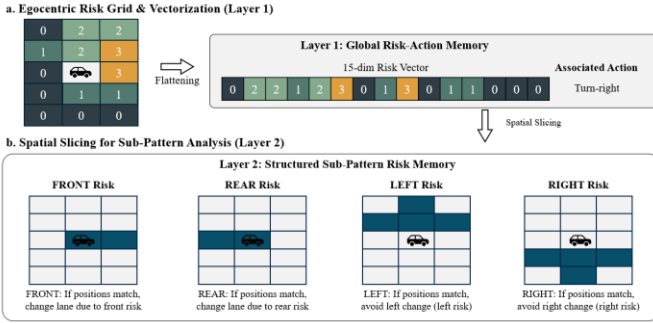


Fig. 3. Two-layer structured memory in RESPOND. Layer 1 stores global risk–action pairs via vectorized risk grids, while Layer 2 decomposes the grid into sub-patterns for tactical reasoning. FRONT/REAR patterns guide strategic actions, and LEFT/RIGHT patterns impose safety constraints.

To support risk-aware decision reuse, RESPOND adopts a two-layer structured memory (Fig. 3). Layer 1 stores global risk–action pairs by flattening the 5×3 grid into a 15-dimensional vector, enabling precise matching for familiar scenarios. To improve generalization, Layer 2 decomposes the grid into semantic sub-patterns. FRONT and REAR patterns capture longitudinal risks for strategic actions, while LEFT and RIGHT patterns represent lateral constraints. This dual structure enables reliable action reuse in high-risk cases and flexible adaptation in novel scenarios.

4.2.1 Layered Memory Architecture

At the core of RESPOND’s memory module is a vectorized representation of the risk pattern defined in Section 4.1. Each 5×3 grid is flattened into a 15-dimensional vector $\mathbf{v} \in \{0,1,2,3\}^{15}$ and paired with an associated action $\mathcal{A}$.

Layer 1: Exact Pattern Memory. The first layer stores tuples $\mathcal{M}_1 = \{(\mathbf{v}_k, \mathcal{A}_k)\}_{k=1}^N$ and supports nearest-neighbor retrieval based on vector similarity. To ensure reliability, action reuse is allowed only when an exact match condition $\|\mathbf{v}_{\text{obs}} - \mathbf{v}_k\| = 0$ is satisfied.

$$(\mathbf{v}^*, \mathcal{A}^*) = \arg \min_{(\mathbf{v}_k, \mathcal{A}_k) \in \mathcal{M}_1} \|\mathbf{v}_{\text{obs}} - \mathbf{v}_k\| \quad (4)$$

where $\mathbf{v}_{\text{obs}}$ is the current observation vector. This vectorized representation removes ambiguity from text-based memory and enables deterministic recall of validated actions. To improve data efficiency, left–right symmetry is exploited by inserting mirrored counterparts of each risk vector.

Layer 2: Sub-Pattern Abstraction Memory. To address sparsity in the combinatorial space of full patterns 4^{15} , this layer stores abstracted sub-patterns that capture critical local configurations. Four canonical types are defined: FRONT and REAR patterns encode longitudinal risks for strategic actions, while LEFT and RIGHT patterns represent lateral constraints. These abstractions enable knowledge transfer across structurally similar but globally distinct scenarios.

4.2.2 Hierarchical Retrieval and Decision Support

During deployment, the memory system performs prioritized retrieval to balance precision and generality:

Exact Reuse (Layer 1): The system first queries Layer 1. If an exact match ($\|\mathbf{v}_{\text{obs}} - \mathbf{v}^*\| = 0$) is found, the corresponding action $\mathcal{A}^*$ is reused directly, ensuring consistent and reliable decisions in previously validated scenarios.

Strategic Inference (Layer 2): If no exact match exists, Layer 2 sub-

patterns are queried to provide high-level intent (e.g., FRONT/REAR for action guidance) or to constrain the action space (e.g., LEFT/RIGHT blocking unsafe maneuvers). These cues are then used by the LLM to generate the final action.

This two-layer design combines compact representation, symmetry-based augmentation, and hierarchical abstraction to enable precise recall, efficient generalization, and scalable learning. It further supports one-crash-to-generalize adaptation by incorporating reflection updates into both memory layers.

4.3. Hybrid Decision-Making Pipeline

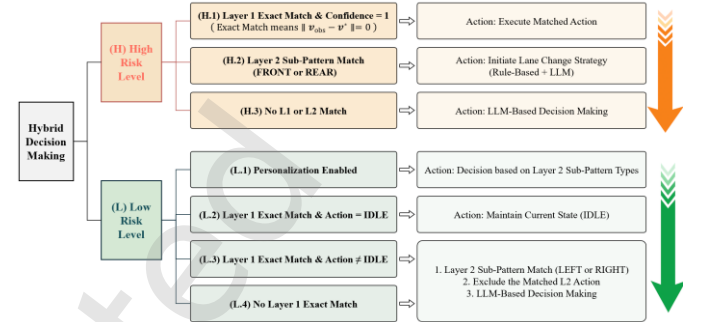


Fig. 4. Hybrid decision-making pipeline in RESPOND. The system first classifies scenarios into high- and low-risk regimes based on DRF-derived risk values, and then applies different strategies, including exact pattern reuse, sub-pattern-guided reasoning, and LLM-based fallback, to ensure safe and adaptive decision-making.

While structured memory enables precise retrieval, autonomous driving requires a decision mechanism that balances safety and adaptability. RESPOND addresses this by adopting a hybrid decision-making pipeline that integrates rule-based constraints with LLM reasoning, and dynamically adjusts strategies based on quantified risk.

4.3.1 Risk Quantification and Regime Classification

The pipeline first quantifies the scene risk. The instantaneous risk level RL_t is defined as the maximum longitudinal risk derived from the unified encoding:

$$RL_t = \max(RV_{\text{front}}, RV_{\text{rear}}) \quad (5)$$

A threshold $\tau = 0.75$ is used to divide the decision process into two regimes: a Safety-Critical regime $RL_t \geq \tau$ for collision avoidance, and an Efficiency-Oriented regime $RL_t < \tau$ for driving comfort and style.

4.3.2 Safety-Critical Regime (High Risk)

In high-risk contexts $RL_t \geq 0.75$, the system follows a three-tier strategy to ensure deterministic safety (Fig. 4). (H.1) Exact Pattern Reuse: The system first queries Layer 1 memory $\mathcal{M}_1$. If an exact match ($\|\mathbf{v}_{\text{obs}} - \mathbf{v}^*\| = 0$) is found, the corresponding action $\mathcal{A}^*$ is executed directly, bypassing LLM reasoning to ensure validated and low-latency responses. (H.2) Sub-Pattern Constrained Reasoning: If no exact match is found, Layer 2 sub-patterns (e.g., FRONT or REAR) are used to activate strategic intent. Rule-based constraints restrict the candidate action space to a safe subset, from which the LLM selects the final action. (H.3) Zero-Shot Fallback: If no memory match exists, the system resorts to LLM-based reasoning using structured risk inputs to generate conservative actions.

4.3.3 Efficiency-Oriented Regime (Low Risk)

In low-risk contexts $RL_t < 0.75$, the system focuses on personalization and efficiency (Fig. 4). (L.1) Personalized Adaptation: If enabled, stylistic sub-patterns from Layer 2 are used to guide action selection according to user preferences. (L.2) Idle Optimization: If a Layer 1 match indicates an IDLE action, it is executed directly, reducing inference cost during steady-state driving. (L.3/L.4) Constraint-Aware Reasoning: For non-idle or novel scenarios, Layer 2 lateral sub-patterns (LEFT/RIGHT) impose negative constraints. If $RV_{lat} > \tau_{lat}$, corresponding lane-change actions are masked from the LLM output space to ensure safety. The lateral masking threshold is set to $\tau_{lat} = 0.67$.

4.4. Pattern-Aware Reflection Mechanism

While the hybrid decision pipeline enables adaptive reasoning, conventional learning methods often require large-scale data to achieve stable performance. RESPOND addresses this limitation through a Pattern-Aware Risk Memory Reflection (RMR), which enables rapid learning from failures by coupling structured risk patterns with symbolic abstraction.

4.4.1 Structured Reflection

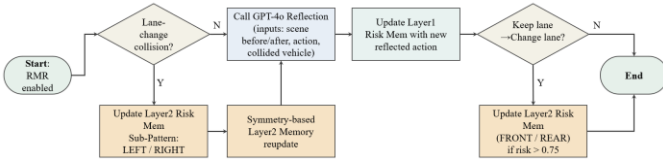


Fig. 5. Pattern-Aware Risk Memory Reflection (RMR). Upon detecting a failure, the system performs structured reflection to identify causes, update memory, and generate corrective actions. Lateral collisions trigger direct sub-pattern updates, while complex cases invoke LLM-based analysis, enabling generalized learning from limited failures.

Reflection is triggered upon detecting a collision or safety violation (Fig. 5). Two types of failures are handled differently to enable targeted updates:

Lane-Change Collisions (Lateral Updates): For collisions caused by lateral maneuvers, the system directly infers spatial constraints and updates LEFT or RIGHT sub-patterns in Layer 2. Symmetry is applied to mirror the constraint, improving learning efficiency.

General Collisions (LLM-Based Analysis): For more complex cases, a high-capacity LLM is invoked with structured inputs to perform causal reasoning and generate a revised action α^* .

4.4.2 Memory Update and Abstraction

The reflected action α^* is written to Layer 1 memory with confidence 1. When failures indicate a higher-level strategy shift (e.g., keep lane to change lane) under high-risk conditions (RV_{front} or $RV_{rear} \geq 0.75$), corresponding sub-patterns (FRONT/REAR) are updated in Layer 2, enabling abstraction from specific failures to generalizable strategies.

The reflection mechanism integrates hierarchical credit assignment, symmetry-based augmentation, and strategy-level abstraction to support both precise correction and generalized learning across scenarios.

Overall, RMR establishes a structured post-failure learning loop that combines causal reasoning with memory updates, enabling one-crash-to-generalize adaptation and continual improvement in

high-risk and long-tail driving scenarios.

4.5. Human-in-the-Loop Personalization

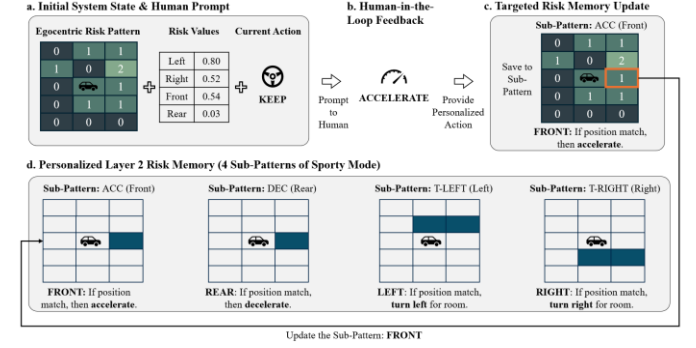


Fig. 6. Human-in-the-loop personalization in RESPOND. The system presents the current risk pattern and action, receives user feedback, and updates Layer 2 sub-pattern memory to encode personalized driving preferences for future reuse.

While reflection enables autonomous improvement from failures, capturing subjective driving preferences requires explicit human guidance. RESPOND addresses this through a human-in-the-loop personalization module built on Layer 2 memory, enabling efficient style adaptation via lightweight feedback without retraining.

4.5.1 Interactive Learning Workflow

The personalization process follows a three-step cycle (Fig. 6). First, the system presents the current context, including the 5×3 risk pattern and associated risk values. Second, the user provides a preferred action. Third, this feedback is abstracted into a stylistic sub-pattern and stored in Layer 2. During deployment, when $RL < 0.75$, these personalized sub-patterns are prioritized to guide action selection in similar contexts.

4.5.2 Efficiency and Safety Constraints

This design follows three principles: (i) safety-constrained adaptation, where personalization is applied only when $RL < 0.75$; (ii) abstract-level learning, which encodes preferences as generalized sub-patterns; and (iii) scalability, enabled by the symbolic Layer 2 representation for cross-scenario transfer.

RESPOND's human-in-the-loop module enables interpretable and efficient personalization by embedding user preferences into the structured sub-pattern space. It supports fine-grained behavior adaptation from sparse feedback while maintaining safety constraints and enabling reusable driving profiles, facilitating scalable deployment across consumer AVs and robotaxi platforms.

5 Experiments

This section evaluates RESPOND through closed-loop simulation, personalization, and real-world experiments. We examine three aspects: (i) safety and generalization enabled by structured risk representation and memory, (ii) learning efficiency from reflection, and (iii) zero-shot transfer to real-world driving data (Huang, Zheng, et al. 2020).

5.1. Experimental Setup and Metrics

We use two platforms: the highway-env simulator (Leurent et al. 2018) for controlled evaluation under varying traffic conditions, and the HighD dataset (Krajewski et al. 2018) for real-world validation.

Implementation Details. Experiments are implemented in

Python 3.9 with PyTorch 2.5. GPT-4o-mini is used for decision-making, and GPT-4o for reflection. In highway-env, we follow the DiLu protocol with 30 decision steps per episode (1 Hz) under varying lane configurations (4-5 lanes) and traffic densities (2.0-3.0).

LLM Configuration. RESPOND uses GPT-4o-mini for decision-making and GPT-4o for reflection. Decoding is deterministic (temperature = 0) with a maximum token limit of 2000. Additional implementation details and prompts are provided in the open-source repository for reproducibility.

Evaluation Metrics. We evaluate performance using: (i) Success Rate (SR), the percentage of collision-free episodes; (ii) Learning Efficiency, the performance gain per reflection iteration; (iii) Risk Reduction ΔR , the decrease in cumulative DRF intensity; and (iv) Human Evaluation Rate, the proportion of expert judgments favoring RESPOND over human actions.

5.2. Comparative Analysis in Closed-Loop Simulation

We evaluate RESPOND in closed-loop simulation using the Highway-env platform under varying traffic densities. Each episode lasts 30 seconds with a 1 Hz decision frequency, following the DiLu protocol.

We compare RESPOND with three baselines: DiLu (LLM-based), GRAD (RL-based), and PADriver (MLLM-based). Evaluations are conducted under three settings: L4-2.0, L5-2.5, and L5-3.0, as summarized in Table 1.

Table 1. Driving Success Rate (SR %) across traffic densities.

Method	L4-2.0	L5-2.5	L5-3.0	Δ (pp)
GRAD	69%	47%	10%	↓ 59
DiLu	70%	65%	35%	↓ 35
RESPOND-L1	85%	75%	50%	↓ 35
RESPOND-L1+L2	100%	80%	70%	↓ 30
RESPOND-Sporty	80%	-	-	-

Notes: Δ denotes the performance drop from L4-2.0 to L5-3.0. L4 and L5 represent 4-lane and 5-lane configurations.

Performance and Robustness. As shown in Table 1, RESPOND-L1+L2 achieves 100% SR in the standard setting, outperforming DiLu (70%) and GRAD (69%). RESPOND-L1 alone reaches 85% SR, indicating the effectiveness of structured risk encoding. Under more complex conditions (L5-3.0), GRAD drops to 10% and DiLu to 35%, while RESPOND-L1+L2 maintains 70% SR with the smallest performance decline. This demonstrates strong robustness and transferability enabled by sub-pattern abstraction.

Learning Efficiency. RESPOND achieves significantly higher sample efficiency: GRAD requires 600,000 RL episodes to reach 69% SR, and DiLu requires ~40 reflection iterations to reach 70% SR, while RESPOND achieves 100% SR with only 6 collision events through structured memory and reflection.

Personalization. RESPOND-Sporty achieves 80% SR with only 20 feedback steps, demonstrating effective personalization with minimal training overhead while maintaining competitive performance.

5.3. Efficiency of Reflection-Based Learning

A key challenge in autonomous driving is learning from sparse failure events. We evaluate RESPOND's reflection efficiency in highway-env under controlled conditions with empty initial memory.

Table 2. Reflection efficiency and generalization performance.

Metric	DiLu	RESPOND-L1	RESPOND-L1+L2
Collisions	≈40	12	6
Memory/Crash	1 (Text)	1.7 (vec)	2.6 (vec+sub)
Efficiency Gain	1 ×	≈3 ×	≈7 ×
Generalization			
Front/Rear	≈ 30%	57%	100%
Left/Right	≈ 30%	57%	97%

Notes: Collisions denote the number of failures required to reach 70% SR. Generalization denotes the percentage of recurring accidents eliminated after a single reflection.

Learning Efficiency. As shown in Table 2, DiLu requires ~40 collisions to reach 70% SR, while RESPOND-L1+L2 achieves 100% SR with only 6 collisions, corresponding to a ~7× improvement. This gain is attributed to symmetry-based augmentation and sub-pattern abstraction, which enable each failure to produce multiple reusable memory units.

One-Shot Generalization. RESPOND-L1+L2 eliminates 100% of longitudinal and 97% of lateral collision recurrences after a single update, outperforming RESPOND-L1 (57%) and DiLu (~30%). This demonstrates that sub-pattern abstraction enables generalization across structurally similar scenarios.

Impact on Performance. The improved sample efficiency translates to stronger closed-loop performance: RESPOND reaches 100% SR after 6 failures, compared to 69% for GRAD after 600,000 episodes and 70% for DiLu after 40 failures. The memory update process remains interpretable, linking each correction to specific risk patterns.

5.4 Personalization under Low-Risk Conditions

RESPOND supports human-in-the-loop personalization under low-risk conditions using Layer 2 sub-pattern memory. Personalization is achieved through lightweight feedback without model retraining. When $RL < 0.75$, user-provided actions are abstracted into sub-patterns and stored for reuse in similar scenarios.

Table 3. Success Rate (SR) and Average Speed comparison under lane-4-density-2.0 setting.

Method	SR (%)	Avg. Speed (m/s)	Notes
DiLu	93	12.85	10 Hz decision freq.; PADriver setting
PADriver Slow	97	18.46	Safety-focused finetuned agent
PADriver Normal	83	20.13	Balanced performance
PADriver Fast	57	24.12	High-speed, unsafe
RESPOND-L1+L2	100	16.90	Hybrid rule-memory reasoning, 1 Hz
RESPOND-Sporty	80	19.39	20-step HITL personalization

Notes: DiLu and PADriver results are obtained at 10 Hz as reported in. The most rigorous comparison for personalization efficiency is between the RESPOND baseline and its Sporty variant, both consistently evaluated at 1 Hz.

We compare RESPOND with PADriver (Slow, Normal, Fast) and DiLu under the lane-4-density-2.0 setting (Table 3).

RESPOND-L1+L2 achieves 100% SR at 1 Hz, outperforming PADriver Slow (97%) despite its higher 10 Hz frequency. The RESPOND Sporty mode, trained with only 20 feedback steps, achieves 80% SR, comparable to PADriver Normal (83%) and significantly higher than PADriver Fast (57%). These results demonstrate effective personalization with minimal supervision.

Driving styles are reflected in average speeds. PADriver Fast achieves the highest speed but suffers from high collision rates. RESPOND Sporty reaches 19.39 m/s, close to PADriver Normal (20.13 m/s), while RESPOND-L1+L2 maintains a lower speed for safety. These results show that RESPOND can reproduce personalized driving styles while preserving safety.

RESPOND achieves personalization without retraining, significantly reducing computational cost compared to prior methods. Personalization is restricted to low-risk scenarios, ensuring safety is preserved. Overall, RESPOND provides efficient and interpretable style adaptation through structured sub-pattern memory.

5.5 Real-World Robustness Validation on HighD Dataset

Simulation environments lack the variability and long-tail cases of real-world driving. To evaluate RESPOND in realistic conditions, we integrate it with the HighD dataset, a large-scale collection of naturalistic highway trajectories. This experiment assesses whether RESPOND can handle unseen safety-critical scenarios through structured risk abstraction without relying on prior memory matches.

Dataset and Scenario Selection. We extract safety-critical lane-change scenarios from the HighD dataset (Fig. 7). Each lane-changing vehicle is treated as the ego agent, and risk is evaluated with respect to the following vehicle after lane entry. A scenario is labeled high-risk if rear TTC falls below 4.0 s after merging. Using this criterion, we obtain 53 scenarios. For each case, we reconstruct the pre-lane-change decision context, encode it into the 5×3 risk pattern, and compute directional DRF-based risk values.

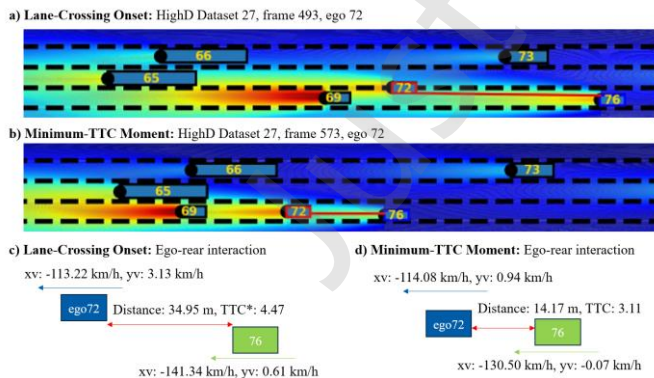


Fig. 7. Example HighD lane-change scenario. (a–b) Global views at lane-crossing onset and minimum-TTC moment. (c–d) Ego-centric interactions with the rear vehicle, showing relative motion, distance, and TTC.

Real-World Intervention Evaluation Protocol. We evaluate whether RESPOND can produce safe decisions in unseen safety-critical scenarios from real-world driving data. The system operates without prior memory matches, relying on structured risk abstraction and LLM reasoning.

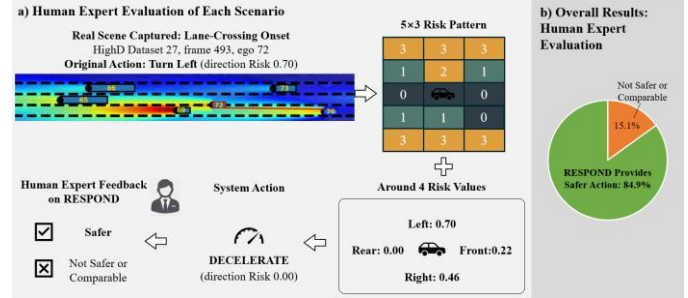


Fig. 8. Human expert evaluation on HighD. (a) Evaluation protocol. (b) Results over 53 scenarios, where RESPOND is judged safer in 84.9% of cases.

For each scenario, RESPOND is queried with structured inputs including spatial configuration, directional risk values, and motion states. The LLM generates an action, which is compared with the recorded human action. Safety is evaluated through both expert judgment (Fig. 8) and a DRF-based quantitative metric. Across 53 scenarios, RESPOND is judged safer in 45 cases, comparable in 5, and worse in 3.

Three independent evaluators reviewed all scenarios using a unified visualization interface. Each labeled RESPOND's action as "Safer", "Comparable", or "Worse". The inter-rater agreement exceeded 0.8 (Cohen's κ), consistent with the quantitative results.

Quantitative Results and Analysis. The results are summarized in Table 4. RESPOND is judged safer in 45 of 53 scenarios (84.9%), comparable in 5 (9.4%), and worse in 3 (5.7%).

It is worth noting that these results are obtained by directly applying the same RESPOND configuration used in simulation, without any additional tuning for real-world data. The prompting strategy remains unchanged, which in some high-speed cut-in scenarios may lead the system to favor acceleration when surrounding vehicles exhibit unusually high velocities.

In addition, real-world traffic includes vehicle characteristics not fully captured in the simulator. The current DRF formulation approximates vehicles using front-point representations, which may introduce inaccuracies for large vehicles such as trucks.

These observations highlight the limitations of directly transferring a simulator-calibrated configuration to real-world data. The purpose of this evaluation is therefore not to demonstrate immediate deployability, but to assess whether RESPOND retains meaningful decision-making capability under real-world conditions.

Table 4. Zero-shot evaluation results on high-risk HighD scenarios.

Comparison Outcome	Count	Percentage
Human Action Safer	3	5.7%
Comparable	5	9.4%
RESPOND Action Safer	45	84.9%
Total Cases	53	100%

Notes: Safer Decision Rate denotes the percentage of cases where RESPOND is judged safer than the recorded human action across 53 high-risk HighD scenarios.

Qualitative analysis shows that RESPOND leverages DRF-based risk patterns to avoid unsafe maneuvers, such as suppressing risky lane changes under limited lateral space and maintaining safe behavior in high-speed interactions. These results

indicate that the structured risk representation provides effective guidance for LLM-based reasoning in unseen scenarios. The zero-shot results on HighD further highlight key architectural strengths. The effective transfer from simulation to real data suggests the robustness of the 5×3 risk representation without additional retraining. In addition, risk-quantified LLM reasoning bridges physical dynamics and symbolic planning, enabling RESPOND to serve as a fallback policy for long-tail scenarios and support offline safety analysis.

5.6 Ablation and Discussion

To evaluate the contribution of key components, we conduct ablation studies on the hierarchical memory, hybrid decision pipeline, and structured risk encoding.

Impact of Layer 2 Sub-Pattern Memory. Removing Layer 2 leads to significant performance degradation, particularly in complex scenarios (Table 5). While RESPOND-L1+L2 maintains strong performance, SR in L5-3.0 drops from 70% to 50% with RESPOND-L1 and further to 35% without memory. This demonstrates that Layer 2 enables strategic generalization by transferring knowledge across structurally similar scenarios, which is essential for learning from sparse failures, while Layer 1 primarily supports precise matching.

Table 5. Effect of memory layers on SR (%) across environments.

Scenario	RESPOND-L1+L2	RESPOND-L1	No Memory
L4-2.0	100.0	85.0	70.0
L5-2.5	80.0	75.0	55.0
L5-3.0	70.0	50.0	35.0

Effect of Hybrid Rule–LLM Decision. Disabling rule-based constraints leads to unsafe lane-change attempts in 20–30% of high-risk cases and a 2× increase in latency. The hybrid design enforces safety constraints while allowing the LLM to focus on high-level reasoning, improving both safety and efficiency.

Utility of Structured Risk Pattern. Compared to textual scene descriptions, the 5×3 risk pattern significantly improves retrieval precision. Text-based memory suffers from ID mismatches and numeric noise, while the structured representation aligns spatial topology with risk, enabling deterministic retrieval and stable action reuse.

Contribution of Risk Values. Removing DRF-based risk inputs significantly degrades performance (Table 6), with the HighD safer rate dropping from 84.9% to 72.6% and L4-2.0 SR from 70% to 50%. Without risk cues, the LLM tends to underestimate braking urgency and overestimate lateral safety, highlighting the importance of quantitative risk priors.

Table 6. Contribution of directional risk values to decision quality.

Metric	With Risk	Without Risk
HighD Safer Rate (%)	84.9	72.6
L4-2.0 Success Rate (%)	70.0	50.0

Role of Exact Retrieval Strategy. Replacing exact matching with approximate nearest-neighbor (ANN) search degrades performance, as geometrically similar vectors may correspond to semantically different patterns. RESPOND therefore separates precision (Layer 1 exact matching) from generalization (Layer 2 abstraction).

Sensitivity to DRF Threshold. We evaluate the robustness of the DRF threshold θ under RESPOND-L1+L2 in the L4-2.0 setting,

with θ varied in [0.005, 0.02].

Table 7. Sensitivity of DRF threshold θ on performance.

θ	Success Rate (%)	Avg. Speed (m/s)
0.005	100.0	16.8
0.01	100.0	16.9
0.02	100.0	17.2

As shown in Table 7, both SR and average speed remain stable under moderate variations of θ , supporting the robustness of the discrete risk representation to threshold perturbations.

In summary, the ablation results validate the design of RESPOND: Layer 2 enables generalization across scenarios, the hybrid pipeline stabilizes decisions, structured risk encoding ensures retrieval precision, and quantitative risk values provide essential safety priors. Together, these components support robust performance in both simulation and real-world long-tail scenarios.

6 Conclusion

This paper presents RESPOND, a knowledge-driven autonomous driving framework that rethinks how LLM-based agents perceive, store, and learn from risk. By encoding ego-centric risk into a discrete 5×3 matrix, RESPOND addresses the limitations of unstructured textual memory and enables precise retrieval and hybrid decision-making, bridging LLM reasoning with safety-critical control. Extensive experiments show that RESPOND achieves 100% success rate in standard highway simulations, outperforming both textual-memory and reinforcement learning baselines, while maintaining robustness under dense and shifted traffic conditions. The pattern-aware reflection mechanism improves learning efficiency by approximately 7×, enabling rapid mitigation of recurring failures. Zero-shot evaluation on the HighD dataset further demonstrates that RESPOND can generate risk-aware decisions in previously unseen safety-critical scenarios. In addition, personalization experiments show that RESPOND adapts efficiently with minimal human feedback. Overall, these results highlight structured risk encoding and symbolic–statistical reasoning as an effective foundation for generalizable and memory-efficient driving intelligence, suggesting RESPOND as a promising framework for safety-critical autonomous driving.

References

Bounini, F., Gingras, D., Pollart, H., Gruyer, D., 2017. Modified artificial potential field method for online path planning applications. In: IEEE Intelligent Vehicles Symposium, 180–185.

Cheng, H., Jiang, Y., Zhang, H., et al., 2025. Emergency index (EI): A two-dimensional surrogate safety measure considering vehicles’ interaction depth. *Transport Res C-Emerg*, 171, 104981.

Cui, C., Ma, Y., Cao, X., Ye, W., Wang, Z., 2024. Receive, reason, and react: Drive as you say, with large language models in autonomous vehicles. *IEEE Intell Transp Syst Mag*, 16(4), 81–94.

Hao, H., Wang, Y., Chen, J., 2024. Empowering scenario planning with artificial intelligence: A perspective on building smart and resilient cities. *Engineering*, 43, 272–283.

Helbing, D., Molnar, P., 1995. Social force model for pedestrian

- dynamics. *Phys Rev E*, 51(5), 4282.
- Hou, X., Wang, W., Yang, L., et al., 2025. Driveagent: Multi-agent structured reasoning with LLM and multimodal sensor fusion for autonomous driving. *arXiv Preprint arXiv:2505.02123*.
- Huang, H., Liu, J., Zhang, B., Zhao, S., Li, B., Wang, J., 2025. LEAD: Learning-enhanced adaptive decision-making for autonomous driving in dynamic environments. *IEEE Trans Intell Transp Syst*.
- Huang, H., Wang, J., Fei, C., et al., 2020. A probabilistic risk assessment framework considering lane-changing behavior interaction. *Sci China Inf Sci*, 63(9), 190203.
- Huang, H., Zheng, X., Yang, Y., Liu, J., Liu, W., Wang, J., 2020. An integrated architecture for intelligence evaluation of automated vehicles. *Accid Anal Prev*, 145, 105681.
- Kolekar, S., De Winter, J., Abbink, D., 2020. Human-like driving behaviour emerges from a risk-based driver model. *Nat Commun*, 11(1), 4850.
- Kou, G., Jia, F., Mao, W., et al., 2025. PADriver: Towards personalized autonomous driving. *arXiv Preprint arXiv:2505.05240*.
- Krajewski, R., Bock, J., Kloeker, L., Eckstein, L., 2018. The HighD dataset: A drone dataset of naturalistic vehicle trajectories on German highways. In: *IEEE Intell Transp Syst Conf*, 2118–2125.
- Lauri, M., Hsu, D., Pajarinen, J., 2022. Partially observable Markov decision processes in robotics: A survey. *IEEE Trans Robot*, 39(1), 21–40.
- Leurent, E., et al., 2018. An environment for autonomous driving decision-making.
- Liu, Y., Wu, F., Liu, Z., Wang, K., Wang, F., Qu, X., 2023. Can language models be used for real-world urban-delivery route optimization? *Innovation*, 4(6).
- Liu, Y., Zhou, Y., Liu, Y., Xu, Z., He, Y., 2025. Intelligent fault diagnosis for CNC through the integration of large language models and domain knowledge graph. *Engineering*.
- Mao, J., Qian, Y., Ye, J., Zhao, H., Wang, Y., 2023. GPT-driver: Learning to drive with GPT. *arXiv Preprint arXiv:2310.01415*.
- Munro, I.C., Renwick, A.G., Danielewska-Nikiel, B., 2008. The threshold of toxicological concern (TTC) in risk assessment. *Toxicol Lett*, 180(2), 151–156.
- Peesapati, L.N., Hunter, M.P., Rodgers, M.O., 2018. Can post encroachment time substitute intersection characteristics in crash prediction models? *J Safety Res*, 66, 205–211.
- Ross, S., Pineau, J., Chaib-draa, B., Kreitmann, P., 2011. A Bayesian approach for learning and planning in partially observable Markov decision processes. *J Mach Learn Res*, 12(5).
- Shalev-Shwartz, S., Shammah, S., Shashua, A., 2017. On a formal model of safe and scalable self-driving cars. *arXiv Preprint arXiv:1708.06374*.
- Spaan, M.T.J., 2012. Partially observable Markov decision processes. In: *Reinforcement Learning: State-of-the-Art*. Springer.
- Wang, J., Huang, H., Li, K., Li, J., 2021. Towards the unified principles for level 5 autonomous vehicles. *Engineering*, 7(9), 1313–1325.
- Wang, J., Wu, J., Zheng, X., Ni, D., Li, K., 2016. Driving safety field theory modeling and its application in pre-collision warning system. *Transport Res C-Emerg*, 72, 306–324.
- Wang, W., Xie, J., Hu, C., et al., 2023. DriveMLM: Aligning multi-modal large language models with behavioral planning states for autonomous driving. *arXiv Preprint arXiv:2312.09245*.
- Wang, Y., Huang, H., Xu, Z., Sun, K., Guo, B., Zhao, J., 2025. Learning from risk: LLM-guided generation of safety-critical scenarios with prior knowledge. *arXiv Preprint arXiv:2511.20726*.
- Wen, L., Yang, X., Fu, D., et al., 2023. DILU: A knowledge-driven approach to autonomous driving with large language models. *arXiv Preprint arXiv:2309.16292*.
- Wen, L., Yang, X., Fu, D., et al., 2023. On the road with GPT-4V: Early explorations of vision-language models for autonomous driving. *arXiv Preprint arXiv:2311.05332*.
- Zeng, F., Wang, S., Zhu, C., Li, L., 2025. ADRD: LLM-driven autonomous driving based on rule-based decision systems. *arXiv Preprint arXiv:2506.14299*.
- Zhao, S., Zhang, J., Masoud, N., Huang, H., Hou, X., He, C., 2025. SACA: A scenario-aware collision avoidance framework for autonomous vehicles integrating LLM-driven reasoning. *arXiv Preprint arXiv:2504.00115*.
- Zhou, Z., Huang, H., Li, B., Zhao, S., Mu, Y., Wang, J., 2026. SafeDrive: Knowledge- and data-driven risk-sensitive decision-making for autonomous vehicles with large language models. *Accid Anal Prev*, 224, 108299.

Author contributions

CRedit authorship contribution statement

Dan Chen: Conceptualization, Formal analysis, Writing – original draft, Visualization. **Heye Huang:** Writing – review & editing, Visualization, Validation. **Tiantian Chen:** Data curation, Investigation. **Zheng Li:** Data curation, Investigation. **Yongji Li:** Formal analysis, Visualization, Investigation. **Yuhui Xu:** Formal analysis, Visualization, Investigation. **Sikai Chen:** Supervision, Conceptualization, Writing – review & editing.

Replication and data sharing

Data will be made available on request.

The source codes and replication package are available on ETS data at <https://doi.org/10.26599/ETSD.2025.9190001>.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Declaration of generative AI and AI-assisted technologies in the writing process (if applicable)

The authors used ChatGPT for language polishing and minor editing. The authors carefully reviewed and revised all content and take full responsibility for the manuscript.

Author biography



learning, autonomous driving, and intelligent agent systems.

Dan Chen received the M.S. degree in Computer Software and Theory from Sun Yat-sen University, Guangzhou, China. He was a visiting scholar at the School of Vehicle and Mobility, Tsinghua University, Beijing, China. His research interests include LLM-based decision-making, risk-aware representation



making and LLM reasoning.

Heye Huang received the Ph.D. degree in mechanical engineering from Tsinghua University, Beijing, China. She is currently a Assistant Professor at the Korea Advanced Institute of Science and Technology. Her research interests include safe and trustworthy autonomy, generative AI, risk-sensitive decision



Yongji Li received the Ph.D. degree in Computer Software and Theory from Sun Yat-sen University, Guangzhou, China. He is currently a software architect and researcher. His research interests include GPU computing, LLM pretraining, and reinforcement learning.



Yuhui Xu received the M.S. degree from Northeastern University, Shenyang, China. He is currently a senior engineer at China Mobile. His research interests include QoS guarantee for high-speed vehicles in 5G networks and UAV path optimization.

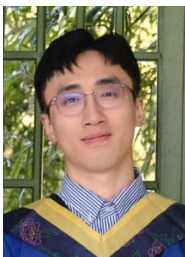


Sikai (Sky) Chen is an Assistant Professor in the Department of Civil and Environmental Engineering and the Department of Mechanical Engineering, by courtesy, at the University of Wisconsin-Madison. He received his Ph.D. degree in civil engineering, with a focus on computational science and engineering, from Purdue University in 2019. His research focuses on human users, AI, and transportation, aiming to develop safe, efficient, sustainable, and human-centered transportation systems.



travel behavior, transport policy, and statistical methods.

Tiantian Chen is an Associate Professor at the Cho Chun Shik Graduate School of Mobility, KAIST. She received her Ph.D. degree in civil and environmental engineering from The Hong Kong Polytechnic University in 2021. Her research focuses on traffic safety, human factors, human-centered design, driving simulation,



His research interests include autonomous vehicle control, evaluation, and simulation-based optimization for traffic systems design.

Zheng Li received the B.S. degree in civil engineering from Hunan University, China, in 2020, and the M.S. degree in transportation engineering from Tongji University, China, in 2023. He is currently a Ph.D. student and Research Assistant in the Department of Civil and Environmental Engineering at the University of Wisconsin-