RETRIEVAL-AUGMENTED LARGE LANGUAGE MODEL DECISION-MAKING FOR AUTONOMOUS DRIVING GUIDED BY CHINESE PHILOSOPHICAL WISDOM

Anonymous authors
Paper under double-blind review

ABSTRACT

Autonomous driving decision systems must balance safety, efficiency, and social norms in complex traffic interactions. Philosophical and ethical considerations have received limited attention in existing autonomous driving decision-making approaches based on numerical optimization, sequence prediction, and large language models (LLMs). We propose Chinese Philosophical Wisdom-Guided Driving (CPW-Drive), a closed-loop retrieval-augmented generation (RAG) framework that incorporates value guidance derived from Chinese philosophy into autonomous driving decision-making. Using Chinese Confucian thought as its knowledge source, CPW-Drive consolidates LLM-extracted keywords from relevant classical texts into driving-relevant value principles through manual screening and validation. It then contextualizes these principles through scenario-specific cases to form retrievable and reusable value guidance. We further propose Physics-aware Spatial Similarity Retrieval (PSSR), which compares vehicle layouts and velocity-extrapolated states to retrieve physically relevant historical cases. On Highway-env's multilane highway-driving task, CPW-Drive achieves success rates of 93.0%, 86.0%, and 72.0% across three traffic configurations. These results outperform the strongest baseline by 8.0, 22.5, and 25.0 percentage points, respectively. Across all configurations, CPW-Drive achieves the highest collision-free step count and maintains a low lane-change frequency. The results suggest that structured value guidance can improve simulated closed-loop safety and stability while introducing efficiency and latency trade-offs.

Keywords: Autonomous driving, Large language models, Chinese philosophical wisdom, Confucian thought, Physics-aware spatial similarity retrieval

1 INTRODUCTION

Recent advances in autonomous driving have improved perception, planning, and control. However, data-driven approaches, including reinforcement learning (RL) and imitation learning (IL), still face interpretability and generalization challenges in open-ended, long-tail scenarios (Kiran et al., 2022; Zhu & Zhao, 2022; Tian et al., 2025a). Moreover, numerical optimization and prediction objectives do not explicitly encode every social norm that can influence human driving. Interactive traffic therefore requires trade-offs among safety, efficiency, and traffic etiquette.

Knowledge-driven autonomous driving has consequently attracted increasing attention. Recent studies have explored large language models (LLMs) for driving decisions because they can combine language-based reasoning, external knowledge, and scene descriptions (Mao et al., 2024; Wen et al., 2024). Most LLM-based driving methods rely on task instructions, commonsense knowledge, or historical experience (Mao et al., 2023; Sha et al., 2023; Wen et al., 2024). More recent systems retrieve traffic regulations or learn value preferences (Cai et al., 2026; Xia et al., 2026). However, the literature has not established how culturally grounded philosophical principles can be represented as reusable, scenario-level driving examples. This gap limits the systematic application of abstract value principles to concrete traffic interactions.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

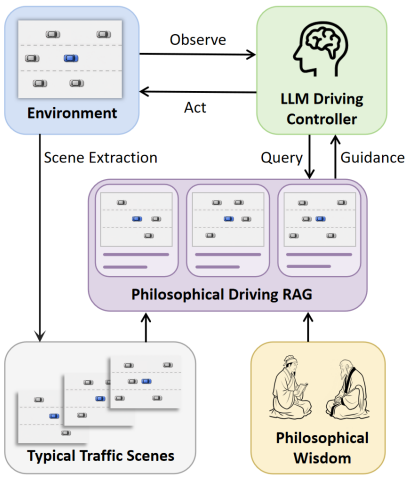

Figure 1: Conceptual illustration of CPW-Drive. Driving-relevant value principles derived from Chinese philosophical wisdom provide retrievable guidance for closed-loop LLM-based driving decisions.

To address this gap, we propose Chinese Philosophical Wisdom-Guided Driving (CPW-Drive), an LLM-based framework that integrates Chinese philosophical wisdom into autonomous driving decision-making. The framework transforms abstract values into structured driving knowledge for concrete traffic scenarios. We select Confucian thought as the knowledge source, extract high-frequency keywords from relevant classical texts, and consolidate them into ten driving-relevant value principles through manual screening and conceptual integration. These value principles are then combined with specific driving scenarios to form retrievable and reusable value-guided driving cases, while Physics-aware Spatial Similarity Retrieval (PSSR) retrieves relevant historical cases for the current decision. Figure 1 illustrates the overall concept.

The main contributions of this paper are summarized as follows:

- To the best of our knowledge, CPW-Drive is the first LLM-based autonomous driving framework to incorporate driving-relevant value principles derived from Chinese philosophical wisdom.
- We derive driving-relevant value principles from Confucian classics and organize them into retrievable, scenario-specific cases, and introduce PSSR for historical scene retrieval.
- Closed-loop experiments across three traffic configurations show higher success rates, more collision-free steps, and restrained driving behavior under Confucian value guidance, with an efficiency trade-off.

2 Related Work

2.1 LLMs for Autonomous Driving Decision-Making

The reasoning and commonsense capabilities of LLMs have extended their use to sequential decision systems, as demonstrated by reasoning-and-acting, self-reflection, and skill-acquisition agents such as ReAct, Reflexion, and Voyager (Yao et al., 2023; Shinn et al., 2023; Wang et al., 2024). LLMs have also been applied to autonomous driving decisions. DriveLikeAHuman organizes decisions using a human-inspired cognitive structure (Fu et al., 2024b). DiLu integrates reasoning, memory, and reflection into a knowledge-driven closed-loop driving framework (Wen et al., 2024). GPT-Driver formulates planning as language-sequence generation (Mao et al., 2023), while DriveGPT4 studies interpretable end-to-end autonomous driving with LLMs (Xu et al., 2024). LanguageMPC connects LLM reasoning to vehicle control through adaptive planning (Sha et al., 2023). LMDrive integrates multimodal sensor information with natural-language instructions for closed-loop driving (Shao et al., 2024). LimSim++ provides a closed-loop evaluation platform (Fu et al., 2024a). More recently, DriveGPT4-V2 further explores enhanced closed-loop autonomous driving with multimodal LLMs (Xu et al., 2025).

These methods primarily address language-based planning, commonsense reasoning, and the use of historical experience. DriveReg and KnowVal additionally incorporate traffic regulations or learned value preferences (Cai et al., 2026; Xia et al., 2026). However, how philosophical principles can be grounded in concrete driving scenarios to support decision-making remains underexplored.

2.2 Ethical and Philosophical Values in Autonomous Driving

Early autonomous-driving ethics mainly focused on collision dilemmas and harm allocation in unavoidable accidents. Bonnefon et al. (2016) identified a public-acceptance paradox for utilitarian autonomous vehicles, while Awad et al. (2018) used the Moral Machine to reveal cross-cultural

differences in ethical preferences. Sui & Sunday Grève (2024) proposed a Confucian moral algorithm for trolley-type harm allocation, rather than LLM-based continuous closed-loop driving. However, dilemma studies provide an insufficient basis for policy and limited guidance for continuous real-world driving, while policy discussions address broader normative considerations (Dewitt et al., 2019; Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659), 2020).

Technical approaches largely formalize traffic rules, safety principles, and ethical requirements as constraints, cost functions, or decision rules. The Responsibility-Sensitive Safety (RSS) model encodes safety responsibilities as proper-response rules (Shalev-Shwartz et al., 2017), while legal-rule models represent traffic regulations and exceptions as structured obligations (Bhuiyan et al., 2023). Although operationally clear, these methods provide limited support for broader value considerations.

LLMs offer an alternative means of incorporating regulations, social norms, and ethical knowledge expressed in natural language, with recent driving studies exploring regulatory compliance and preference alignment (Cai et al., 2026; Xia et al., 2026). Beyond driving, ConfuciusLM applies Confucian pedagogical principles through fine-tuning for primary education, whereas CyberConfucius retrieves classical Confucian texts for philosophical counseling (Xu et al., 2026; Zhao et al., 2025). FLAMES and CMoralEval benchmark Chinese values and moral norms for LLM evaluation (Huang et al., 2024; Yu et al., 2024).

Existing studies have incorporated philosophical and culturally situated values into LLM training, retrieval, or evaluation, but their use in continuous driving decision-making remains underexplored. Rather than formalizing Confucian values as scalar rewards or optimization objectives, CPW-Drive represents them through natural-language principles and scenario-grounded historical cases to guide LLM-based closed-loop driving decisions at inference time.

2.3 Retrieval-Augmented Decision-Making and Scene Similarity Matching

In autonomous driving, retrieval-augmented methods use external knowledge or historical experience to support decision-making. Existing approaches commonly retrieve relevant knowledge or historical cases based on natural-language descriptions or embedding representations (Cai et al., 2026; Xia et al., 2026; Ye et al., 2025; Wen et al., 2024; Mao et al., 2024). However, semantically similar representations may correspond to driving scenes with substantially different spatial configurations and motion states.

To better characterize driving scenes themselves, recent studies have incorporated structured driving information, including BEV representations, scene graphs, trajectory graphs, and spatiotemporal structures, into scene retrieval (Tang et al., 2025; Tian et al., 2025b; Cheng et al., 2024; Zhao et al., 2026). These methods typically measure scene similarity through learned scene representations or structural matching, allowing spatial relations, interactions, and trajectory information to be reflected in the retrieval process. In such approaches, scene similarity is generally determined through the resulting representations or structures rather than by an explicit similarity function defined directly over vehicle physical states.

Unlike the representation- or structure-based matching approaches above, PSSR defines scene similarity directly from vehicle physical states. It jointly considers the current spatial configuration and a short-term future spatial state estimated from current vehicle velocities, allowing the similarity calculation to reflect both spatial layout and motion trends.

3 Methodology

Figure 2 illustrates the overall architecture of CPW-Drive, which consists of four main components: driving-relevant value principle construction, philosophical driving knowledge-base construction, historical scene retrieval, and closed-loop LLM decision-making.

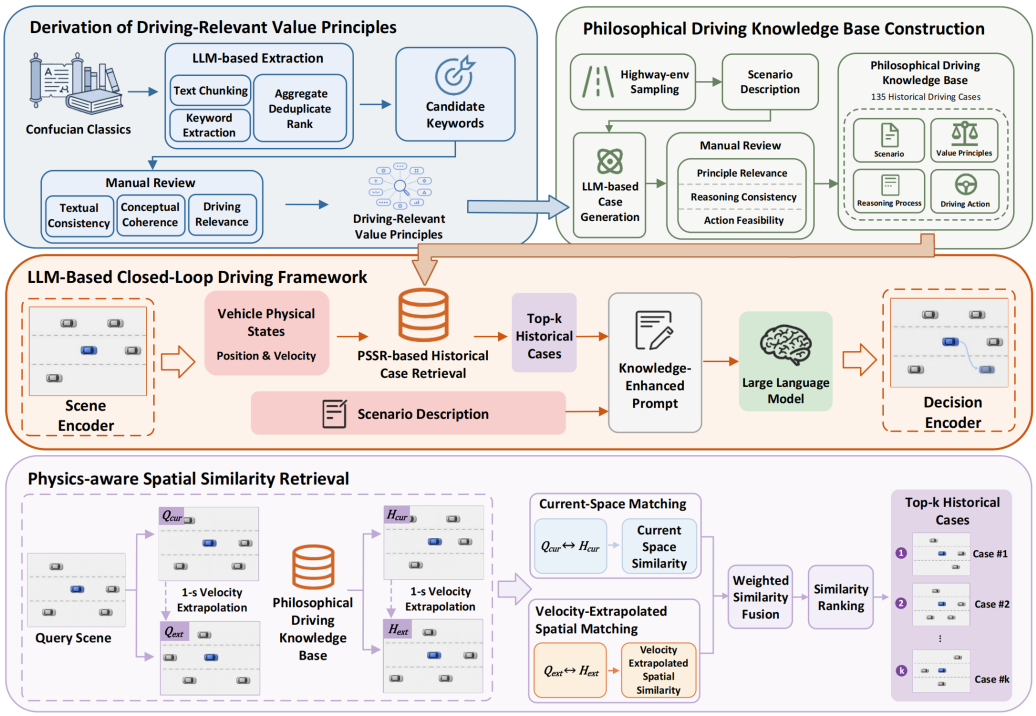

Figure 2: System overview of CPW-Drive. Driving-relevant value principles are extracted and combined with scenario-specific driving cases to construct the Philosophical Driving Knowledge Base, from which PSSR retrieves relevant historical cases to support closed-loop LLM decision-making.

3.1 LLM-Based Closed-Loop Driving Framework

CPW-Drive operates as a continuous observation, decision-making, and execution loop. At each step, the Scene Encoder converts the environment state into a structured natural-language Scenario Description for LLM reasoning. The system retains vehicle physical states for PSSR to retrieve relevant historical cases from the Philosophical Driving Knowledge Base. Each case contains four fields: scenario, value principles, reasoning process, and driving action.

The LLM reasons over the current Scenario Description, retrieved cases, and available driving actions to generate the current reasoning process and driving action. The Decision Encoder processes the generated decision for execution in the driving environment, and the updated environment state enters the next step.

The prompts used by CPW-Drive and an illustrative closed-loop decision example are provided in Appendix B.

3.2 Derivation of Driving-Relevant Value Principles

We use 131 Confucian classics as the knowledge source and divide the texts into chunks, from which an LLM extracts keywords that capture their core ideas. Keywords extracted across different classics and text chunks are then aggregated, deduplicated, and ranked by identification frequency to form a candidate keyword set. The resulting candidate set contained 40 keywords. The titles of the 131 Confucian classics are listed in Appendix A. The keyword-extraction procedure and results are available in the anonymous repository.

Based on the candidate keyword set, we manually review, screen, and conceptually integrate the keywords according to their consistency with the source texts, conceptual overlap, and relevance to driving decision-making, consolidating similar or redundant terms and removing weakly relevant candidates. The resulting set is then subject to a final review by one of the authors, who holds a doctoral degree in Chinese philosophy. This process yields ten driving-relevant value principles:

Wisdom, Righteousness, Respect, Adaptability, Benevolence, Courage, Moderation, Propriety, Empathy, and Trustworthiness.

These value principles provide high-level value considerations rather than directly prescribing driving actions. We further combine them with driving scenarios, reasoning processes, and actions to construct historical driving cases for the philosophical driving knowledge base.

3.3 Philosophical Driving Knowledge Base Construction

The Philosophical Driving Knowledge Base is organized as a set of historical driving cases. Each case contains four fields: scenario, value principles, reasoning process, and driving action, and can be represented as $C = (S, V, R, A)$. Here, S describes the traffic state at the decision moment, including the positions, velocities, and accelerations of the ego vehicle and surrounding vehicles; V denotes the value principles relevant to the scenario; R records the corresponding reasoning process; and A specifies the resulting driving action. This structure preserves both the driving decision and its value-guided reasoning process.

To construct historical driving cases, we collect more than 3,000 candidate scenarios from Highway-env under different lane configurations, vehicle densities, and random seeds, and remove scenarios with highly similar local traffic configurations. For each retained scenario, the scenario information and the full set of value principles derived in the previous section are provided to the LLM, which identifies the relevant principles and generates the corresponding reasoning process and driving action, forming an initial historical case.

The generated cases are then manually reviewed for the appropriateness of the selected value principles, the consistency of the reasoning process with the scenario and selected principles, and the plausibility and feasibility of the driving action. Cases with correctable issues are manually revised, while unsuitable cases are removed. The final knowledge base contains 135 historical driving cases for subsequent PSSR-based retrieval.

3.4 Physics-aware Spatial Similarity Retrieval

PSSR is inspired by Smooth Overlap of Atomic Positions (SOAP), which was originally developed for comparing local atomic structures (Bartók et al., 2013). We adapt its continuous-density-overlap idea to ego-centered driving scenes, introduce anisotropic longitudinal and lateral spatial scales, and estimate short-term future vehicle positions from current velocities to capture both spatial structure and motion trends.

For the current scene A and a historical scene B, both are represented in an ego-centered coordinate system as

$$Z = \left(v^Z_{\mathrm{ego},x}, \left\{ (x^Z_i, y^Z_i, v^Z_{i,x}) \right\}_{i=1}^{N_Z} \right), \qquad Z \in \{A, B\}.$$

Here, $Z \in \{A, B\}$ indexes the current or historical scene, x and y denote the longitudinal and lateral positions of surrounding vehicles relative to the ego vehicle, respectively, v_x denotes the longitudinal vehicle velocity, and $v_{\mathrm{ego},x}$ denotes the ego-vehicle longitudinal velocity. The ego vehicle serves as the coordinate reference and is not included in the subsequent spatial density field, while its velocity is used to characterize the relative motion of surrounding vehicles. All historical scenes in the knowledge base contain surrounding vehicles. When no surrounding vehicle is present in the current scene, PSSR-based historical case retrieval is skipped; hence, the following similarity computation is defined only for non-empty vehicle sets.

In addition to the current spatial state, PSSR estimates a short-term future spatial state from the current vehicle velocities. For vehicle i in scene A, the extrapolated position is defined as

$$\hat{x}^A_i = x^A_i + \left(v^A_{i,x} - v^A_{\mathrm{ego},x} \right) \Delta t, \qquad \hat{y}^A_i = y^A_i, \tag{1}$$

where $\Delta t = 1.0\,\mathrm{s}$. The same extrapolation is applied to scene B. Each scene therefore forms a current-space point set and a velocity-extrapolated point set:

$$(P_{\text{cur}}(A), P_{\text{ext}}(A)) = \left(\left\{ (x_i^A, y_i^A) \right\}_{i=1}^{N_A}, \left\{ (\hat{x}_i^A, \hat{y}_i^A) \right\}_{i=1}^{N_A} \right),$$

with analogous definitions for scene B. By considering both the current vehicle layout and the short-term extrapolated state, PSSR captures spatial configuration together with motion trends in the subsequent similarity calculation. Compared with using the current spatial state alone, incorporating the extrapolated state accounts for relative velocity differences and can therefore retrieve cases with more closely matched motion conditions.

To characterize structural similarity between vehicles in road space, PSSR represents the position of each surrounding vehicle using a two-dimensional anisotropic Gaussian density:

$$\rho_i(x, y) = \exp\left(-\frac{(x - x_i)^2}{2\sigma_x^2} - \frac{(y - y_i)^2}{2\sigma_y^2} \right). \tag{2}$$

Here, σ_x and σ_y control the longitudinal and lateral spatial scales, respectively. Anisotropic spatial scales are used because longitudinal following distances and lateral lane offsets have different physical meanings and scales. The overlap between the Gaussian densities of two vehicles is rescaled by the common factor $\pi\sigma_x\sigma_y$ to define the vehicle-level spatial kernel:

$$K_{\text{space}}(i, j) = \frac{1}{\pi\sigma_x\sigma_y} \iint \rho_i(x, y)\rho_j(x, y)\, dxdy = \exp\left(-\frac{(x_i - x_j)^2}{4\sigma_x^2} - \frac{(y_i - y_j)^2}{4\sigma_y^2} \right). \tag{3}$$

We set $\sigma_x = 5\,\text{m}$, and determine $\sigma_y \approx 1.1533\,\text{m}$ by requiring the spatial kernel to decay to 0.10 at a lateral offset of one lane width, $L = 3.5\,\text{m}$.

For two vehicle point sets X and Y, the scene-level raw spatial overlap is defined as the sum of the spatial similarities over all vehicle pairs:

$$S_{\text{space}}(X, Y) = \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} \exp\left(-\frac{(x_i^X - x_j^Y)^2}{4\sigma_x^2} - \frac{(y_i^X - y_j^Y)^2}{4\sigma_y^2} \right). \tag{4}$$

Because the computation sums over all vehicle pairs across the two scenes, it does not require an explicit one-to-one vehicle correspondence. To enable comparisons between scenes with different numbers of vehicles, we use the normalized form

$$K_{\text{space}}^0(X, Y) = \frac{S_{\text{space}}(X, Y)}{\sqrt{S_{\text{space}}(X, X) S_{\text{space}}(Y, Y)}}. \tag{5}$$

Here, $S_{\text{space}}(X, X)$ denotes the self-overlap of scene X, which includes cross terms between distinct vehicles within the scene and therefore cannot be simplified to the number of vehicles.

The current-space similarity and velocity-extrapolated-space similarity can be written compactly as

$$K_s(A, B) = K_{\text{space}}^0\left(P_s(A), P_s(B)\right), \qquad s \in \{\text{cur}, \text{ext}\}. \tag{6}$$

The two similarities are combined through a weighted average:

$$K(A, B) = w_{\text{cur}} K_{\text{cur}}(A, B) + w_{\text{ext}} K_{\text{ext}}(A, B). \tag{7}$$

We set $w_{\text{cur}} = 0.7$ and $w_{\text{ext}} = 0.3$, giving greater weight to the current spatial layout while using the extrapolated state to complement short-term motion trends. Because the available samples for weight tuning are limited, we do not perform an extensive hyperparameter search and instead adopt this heuristic setting. The resulting similarity lies in $(0, 1]$. Detailed derivations of the Gaussian-overlap kernel, the lateral-scale parameter, the normalized similarity, and its value range are provided in Appendix F.

After obtaining the final PSSR similarity $K(A, B)$ between the current scene A and each historical scene B in the knowledge base, we rank the historical cases in descending order of similarity and select the Top-k most relevant cases as references for the current LLM decision, with $k = 3$ in the main experiments. A qualitative case comparison of historical scenes retrieved by text-vector retrieval and PSSR is provided in Appendix D. The computational complexity, working-memory cost, and empirical retrieval-latency comparison with vector-based retrieval methods are provided in Appendix E.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

We evaluate CPW-Drive in the `highway-v0` environment of the Highway-env simulator (Leurent, 2018). The discrete meta-action space contains five driving actions: change lane left, maintain the current state, change lane right, accelerate, and decelerate. Following the experimental settings of DiLu (Wen et al., 2024), we use three traffic configurations with increasing traffic complexity: four lanes with density 2.0, five lanes with density 2.5, and five lanes with density 3.0.

For each configuration, we conduct 10 evaluation rounds, each comprising 20 runs with different random seeds. Each run contains 30 decision steps at a simulated policy frequency of 1 Hz, and a trial is considered successful if the ego vehicle completes all 30 steps without a collision. Main and ablation results are reported as the mean and standard deviation across the 10 rounds. Knowledge-base and test scenarios are generated with non-overlapping random seeds. All main experiments use 135 human-screened historical cases, with the top three PSSR-ranked cases retrieved at each decision step as reasoning demonstrations; wall-clock inference latency is evaluated separately in Appendix E.

We compare CPW-Drive with DriveLikeAHuman (Fu et al., 2024b), DiLu (Wen et al., 2024), and LimSim++ (Fu et al., 2024a). DriveLikeAHuman represents an LLM-based human-like driving decision method, while DiLu and LimSim++ represent LLM-based autonomous driving methods incorporating knowledge retrieval or simulation-based reasoning. All methods use DeepSeek-V4-Pro through the same API without additional fine-tuning. The baselines are reproduced from their publicly available implementations while preserving their original prompt templates and method-specific settings.

We use four evaluation metrics: success rate (SR), collision-free steps (CFS), efficiency score (ES) (Jiang et al., 2025), and lane-change frequency (LCF). SR is the proportion of trials that complete all 30 decision steps without collision, and CFS is the number of steps completed before collision or termination. Following KoMA, ES measures driving efficiency based on the ratio of the ego vehicle's speed to the average speed of surrounding vehicles; it is set to 10 when the ego vehicle's speed reaches or exceeds that average and is otherwise scaled proportionally. LCF records the frequency of lateral maneuvers and is treated as a descriptive behavioral metric rather than a metric with a monotonic preference.

4.2 MAIN PERFORMANCE ANALYSIS

Table 1 shows that CPW-Drive achieves the highest SR and CFS across all three traffic configurations. Compared with the strongest SR baseline in each configuration, its SR improves by 8.0, 22.5, and 25.0 percentage points, respectively. Its LCF also remains generally low, while its ES is lower than the best baseline in all three configurations. Overall, these results exhibit a performance pattern that places greater emphasis on safety and restraint, which is also consistent with the more restrained and accommodating driving tendency reflected in Confucian thought. A complementary third-party evaluation of the Confucian value conformity of CPW-Drive's reasoning and decisions is provided in Appendix C. The following ablation study further examines the roles of retrieval, philosophical knowledge, and PSSR, and tests whether this performance pattern can be explained by generic conservative driving prompts.

Table 1: Main closed-loop driving performance under different traffic settings. Best and second-best results for SR, CFS, and ES within each setting are shown in bold and underlined, respectively. LCF is reported as a descriptive behavioral metric and is not ranked.

(a) Success and safety metrics

	4 lanes, density = 2.0		5 lanes, density = 2.5		5 lanes, density = 3.0	
Model	SR ↑ (%)	CFS ↑	SR ↑ (%)	CFS ↑	SR ↑ (%)	CFS ↑
DriveLikeAHuman	51.5 ± 5.8	19.75 ± 1.25	50.5 ± 10.1	$\underline{24.50 \pm 1.07}$	33.5 ± 5.8	12.90 ± 1.74
DiLu	75.0 ± 7.5	22.90 ± 2.66	59.0 ± 5.7	22.15 ± 1.34	$\underline{47.0 \pm 5.9}$	14.24 ± 1.44
LimSim++	$\underline{85.0 \pm 8.8}$	$\underline{27.35 \pm 1.98}$	$\underline{63.5 \pm 8.8}$	21.25 ± 2.13	37.5 ± 7.9	$\underline{14.32 \pm 2.51}$
CPW-Drive (Ours)	$\mathbf{93.0 \pm 5.4}$	$\mathbf{28.12 \pm 1.51}$	$\mathbf{86.0 \pm 7.0}$	$\mathbf{27.10 \pm 1.55}$	$\mathbf{72.0 \pm 6.3}$	$\mathbf{22.44 \pm 1.37}$

(b) Efficiency and Driving Behavior Metrics

	4 lanes, density = 2.0		5 lanes, density = 2.5		5 lanes, density = 3.0	
Model	ES ↑	LCF (%)	ES ↑	LCF (%)	ES ↑	LCF (%)
DriveLikeAHuman	$\underline{9.38 \pm 0.16}$	14.73 ± 1.74	$\mathbf{9.45 \pm 0.20}$	9.54 ± 1.68	$\underline{9.20 \pm 0.18}$	11.86 ± 2.13
DiLu	9.31 ± 0.17	4.29 ± 0.94	9.27 ± 0.16	5.73 ± 0.73	9.17 ± 0.20	6.58 ± 1.06
LimSim++	$\mathbf{9.49 \pm 0.11}$	8.38 ± 1.43	$\underline{9.42 \pm 0.14}$	8.21 ± 1.27	$\mathbf{9.21 \pm 0.14}$	8.41 ± 0.98
CPW-Drive (Ours)	8.38 ± 0.13	4.93 ± 1.68	8.32 ± 0.11	5.49 ± 0.69	8.47 ± 0.25	6.75 ± 1.24

Table 2: Ablation results under different traffic settings. Best and second-best results for SR, CFS, and ES within each setting are shown in bold and underlined, respectively. LCF is reported as a descriptive behavioral metric and is not ranked.

(a) Success and safety metrics

	4 lanes, density = 2.0		5 lanes, density = 2.5		5 lanes, density = 3.0	
Method	SR ↑ (%)	CFS ↑	SR ↑ (%)	CFS ↑	SR ↑ (%)	CFS ↑
CPW-Drive	$\mathbf{93.0 \pm 5.4}$	$\underline{28.12 \pm 1.51}$	$\mathbf{86.0 \pm 7.0}$	$\mathbf{27.10 \pm 1.55}$	$\mathbf{72.0 \pm 6.3}$	$\mathbf{22.44 \pm 1.37}$
No RAG	84.0 ± 4.6	25.85 ± 1.22	70.0 ± 5.8	24.10 ± 1.27	57.0 ± 6.7	18.05 ± 1.69
Text RAG	84.5 ± 6.0	25.55 ± 1.82	74.5 ± 5.5	24.70 ± 1.45	61.0 ± 7.7	18.60 ± 1.91
RAG w/o Philosophy	85.5 ± 5.0	26.45 ± 1.08	77.0 ± 6.3	25.25 ± 1.32	67.0 ± 5.9	20.35 ± 1.78
Philosophy in Prompt	$\underline{92.0 \pm 4.8}$	27.82 ± 1.57	$\underline{82.0 \pm 5.4}$	$\underline{26.40 \pm 1.27}$	$\underline{69.5 \pm 6.4}$	$\underline{21.05 \pm 1.67}$
Conservative Prompt	$\mathbf{93.0 \pm 2.58}$	$\mathbf{28.33 \pm 0.64}$	76.0 ± 9.07	23.86 ± 3.09	55.0 ± 11.06	18.88 ± 2.98

(b) Efficiency and Driving Behavior Metrics

	4 lanes, density = 2.0		5 lanes, density = 2.5		5 lanes, density = 3.0	
Method	ES ↑	LCF (%)	ES ↑	LCF (%)	ES ↑	LCF (%)
CPW-Drive	8.38 ± 0.13	4.93 ± 1.68	8.32 ± 0.11	5.49 ± 0.69	8.47 ± 0.25	6.75 ± 1.24
No RAG	8.43 ± 0.21	7.55 ± 1.76	8.57 ± 0.31	7.70 ± 1.49	8.48 ± 0.25	9.35 ± 1.84
Text RAG	$\underline{8.67 \pm 0.09}$	7.15 ± 1.61	$\underline{8.63 \pm 0.19}$	8.35 ± 1.42	$\underline{8.83 \pm 0.19}$	8.95 ± 1.65
RAG w/o Philosophy	$\mathbf{8.85 \pm 0.23}$	8.60 ± 1.87	$\mathbf{9.02 \pm 0.10}$	9.75 ± 1.66	$\mathbf{9.08 \pm 0.13}$	10.35 ± 1.92
Philosophy in Prompt	7.54 ± 0.15	4.35 ± 1.29	7.63 ± 0.21	4.70 ± 1.03	7.65 ± 0.16	6.10 ± 1.27
Conservative Prompt	5.72 ± 0.26	1.28 ± 0.16	5.75 ± 0.36	1.80 ± 0.17	6.20 ± 0.38	2.73 ± 0.50

4.3 Ablation Study

We evaluate three ablations and two prompt-based controls to isolate retrieval, PSSR, philosophical content, and generic conservatism. **No RAG** removes historical-case retrieval; **Text RAG** replaces PSSR with text-vector retrieval using the same database and retrieval count; and **RAG w/o Philosophy** retains PSSR and the same historical scenarios but excludes value principles, generating reasoning and actions solely from scene information. **Philosophy in Prompt** directly places the ten value principles in the system prompt without retrieval, whereas **Conservative Prompt** combines non-philosophical historical cases with generic conservative-driving instructions. Both control prompts are provided in Appendix B.

Retrieval ablations support complementary roles for historical cases and PSSR. Relative to CPW-Drive, **No RAG** reduces SR and CFS and increases LCF, showing that retrieved cases support

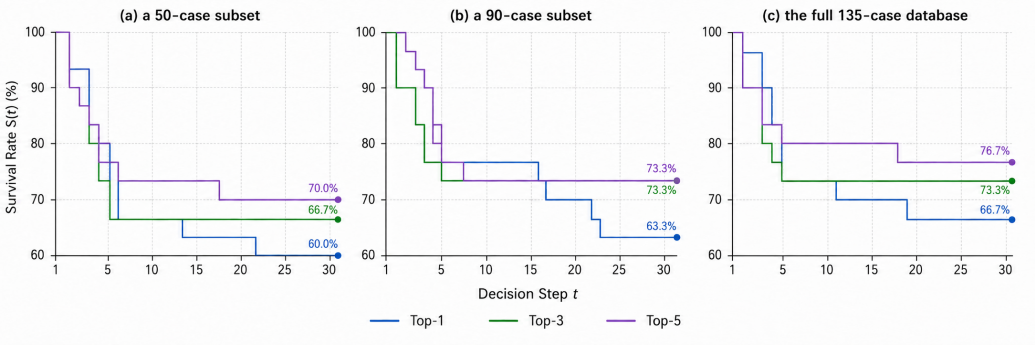

Figure 3: Safe-step survival curves under different database sizes and Top-k settings. Panels (a)–(c) correspond to 50, 90, and 135 historical cases, respectively.

current-scene decisions. **Text RAG** also reduces SR and CFS, indicating that PSSR retrieves better-matched cases by modeling spatial relations and motion states.

The philosophy-related comparisons reveal a safety-efficiency trade-off. Relative to CPW-Drive, **RAG w/o Philosophy** yields lower SR and CFS but higher ES and LCF across all configurations, indicating that philosophical content contributes to safer, more restrained behavior at an efficiency cost. **Philosophy in Prompt** approaches the full model in SR and CFS but further lowers ES and LCF, indicating greater conservatism from direct value injection. Grounding values in concrete scenarios, reasoning processes, and actions clarifies their scene-specific application, mitigating this conservatism and efficiency loss. **Conservative Prompt** produces the lowest LCF but incurs greater efficiency loss and degrades more sharply as traffic complexity increases. Thus, CPW-Drive's performance cannot be explained by generic conservatism alone, and excessive restrictions on proactive maneuvers, such as lane changes and acceleration, may reduce adaptability to complex traffic.

To examine retrieval-database size and top-k settings, we conduct a sensitivity analysis in the highest-density environment. Figure 3 reports 30 random seeds, with at most 30 decision steps per run. We compare databases containing 50, 90, or all 135 historical cases under top-1, top-3, and top-5 retrieval.

The sensitivity results show that overall survival performance improves as the number of historical cases increases, indicating that broader scenario coverage helps retrieve more relevant reference cases. Performance also generally improves as Top-k increases; however, while Top-3 and Top-5 achieve similar final performance, Top-5 exhibits a higher survival rate at earlier stages.

5 Conclusion

This paper presents CPW-Drive, a closed-loop RAG framework that integrates Chinese philosophical wisdom into LLM-based autonomous driving decision-making. The framework derives driving-relevant value principles from Confucian classics and organizes them together with specific driving scenarios, reasoning processes, and actions in the Philosophical Driving Knowledge Base, while PSSR retrieves relevant historical cases according to vehicle spatial structure and motion trends. Across three simulated traffic configurations, CPW-Drive achieves the highest SR and CFS and maintains a relatively low LCF, while exhibiting lower ES.

The current study is limited by a knowledge base containing 135 historical cases and relatively short 30-step trials. In addition, the approximately 20 s mean end-to-end decision latency does not yet meet real-time control requirements. Future work will further evaluate the framework in more diverse driving environments and with larger knowledge bases, while improving its real-time decision-making capability.

AI USE STATEMENT

In this work, generative AI tools were used both as components of the research methodology and as research assistants. DeepSeek-V4-Pro was used to extract candidate keywords from Confucian classical texts, generate initial scenario-specific driving cases, and produce driving decisions in the closed-loop experiments, as described in the methodology and experimental setup. Generative AI tools were also used to support research ideation and execution, assist in searching for potentially relevant prior work, and improve the grammar, clarity, and LaTeX presentation of the manuscript. The authors independently evaluated, selected, and cited all prior work. The extracted value principles and generated driving cases were manually screened, revised, and validated by the authors. All AI-assisted research suggestions and manuscript edits were reviewed and revised where necessary. The authors made all final methodological and interpretive decisions and take full responsibility for the text, claims, results, and artifacts presented in this work.

REPRODUCIBILITY STATEMENT

Appendix B provides the system and user prompt templates, a representative retrieved historical case, and an example LLM decision response. The anonymous repository at `https://anonymous.4open.science/r/CPW-Driver-F8B9` contains the implementations of CPW-Drive and PSSR, the complete prompt materials and historical-case database, simulator and environment configurations, experimental protocols, parameter settings, and scripts for running and evaluating the experiments.

REFERENCES

Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. The Moral Machine Experiment. *Nature*, 563(7729): 59–64, 2018. doi: 10.1038/s41586-018-0637-6.

A. P. Bartók, R. Kondor, and G. Csányi. On Representing Chemical Environments. *Physical Review B*, 87(18):184115, 2013. doi: 10.1103/PhysRevB.87.184115.

Hanif Bhuiyan, Guido Governatori, Andry Rakotonirainy, Meng Weng Wong, and Avishkar Mahajan. Driving Decision Making of Autonomous Vehicle According to Queensland Overtaking Traffic Rules. *The Review of Socionetwork Strategies*, 17(2):233–254, 2023. doi: 10.1007/s12626-023-00147-x.

Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan. The Social Dilemma of Autonomous Vehicles. *Science*, 352(6293):1573–1576, 2016. doi: 10.1126/science.aaf2654.

Tianhui Cai, Yifan Liu, Zewei Zhou, Haoxuan Ma, Seth Z Zhao, Zhiwen Wu, Xu Han, Zhiyu Huang, and Jiaqi Ma. Driving with Regulation: Trustworthy and Interpretable Decision-Making for Autonomous Driving with Retrieval-Augmented Reasoning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(45):38287–38295, 2026. doi: 10.1609/aaai.v40i45.41168.

Ming Cheng, Bowen Zhang, Ziyu Wang, Ziyi Zhou, Weiqi Feng, Yi Lyu, and Xingjian Diao. VeTraSS: Vehicle Trajectory Similarity Search Through Graph Modeling and Representation Learning. *arXiv preprint arXiv:2404.08021*, 2024. doi: 10.48550/arXiv.2404.08021. URL `https://arxiv.org/abs/2404.08021`.

Barry Dewitt, Baruch Fischhoff, and Nils-Eric Sahlin. 'Moral Machine' Experiment Is No Basis for Policymaking. *Nature*, 567(7746):31, 2019. doi: 10.1038/d41586-019-00766-x.

Daocheng Fu, Wenjie Lei, Licheng Wen, Pinlong Cai, Song Mao, Min Dou, Botian Shi, and Yu Qiao. LimSim++: A Closed-Loop Platform for Deploying Multimodal LLMs in Autonomous Driving. In *2024 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1084–1090. IEEE, 2024a. doi: 10.1109/IV55156.2024.10588848.

Daocheng Fu, Xin Li, Licheng Wen, Min Dou, Pinlong Cai, Botian Shi, and Yu Qiao. Drive Like a Human: Rethinking Autonomous Driving with Large Language Models. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, pp. 910–919. IEEE, 2024b. doi: 10.1109/WACVW60836.2024.00102.

Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659). Ethics of Connected and Automated Vehicles: Recommendations on Road Safety, Privacy, Fairness, Explainability and Responsibility. Technical report, Publications Office of the European Union, Luxembourg, 2020. URL https://op.europa.eu/en/publication-detail/-/publication/89624e2c-f98c-11ea-b44f-01aa75ed71a1/language-en/.

Kexin Huang, Xiangyang Liu, Qianyu Guo, Tianxiang Sun, Jiawei Sun, Yaru Wang, Zeyang Zhou, Yixu Wang, Yan Teng, Xipeng Qiu, Yingchun Wang, and Dahua Lin. FLAMES: Benchmarking Value Alignment of LLMs in Chinese. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4551–4591, 2024. doi: 10.18653/v1/2024.naacl-long.256.

Han Jiang, Dongyao Zhu, Xiaoyuan Yi, Ziang Xiao, Zhihua Wei, and Xing Xie. PICACO: Pluralistic In-Context Value Alignment of LLMs via Total Correlation Optimization. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2507.16679.

Kemou Jiang, Xuan Cai, Zhiyong Cui, Aoyong Li, Yilong Ren, Haiyang Yu, Hao Yang, Daocheng Fu, Licheng Wen, and Pinlong Cai. KoMA: Knowledge-Driven Multi-Agent Framework for Autonomous Driving with Large Language Models. *IEEE Transactions on Intelligent Vehicles*, 10(10):4655–4668, 2025. doi: 10.1109/TIV.2024.3488793.

B Ravi Kiran, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A Al Sallab, Senthil Yogamani, and Patrick Pérez. Deep Reinforcement Learning for Autonomous Driving: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):4909–4926, 2022. doi: 10.1109/TITS.2021.3054625.

Edouard Leurent. An Environment for Autonomous Driving Decision-Making. https://github.com/eleurent/highway-env, 2018.

Jiageng Mao, Yuxi Qian, Junjie Ye, Hang Zhao, and Yue Wang. GPT-Driver: Learning to Drive with GPT. *arXiv preprint arXiv:2310.01415*, 2023. doi: 10.48550/arXiv.2310.01415. URL https://arxiv.org/abs/2310.01415.

Jiageng Mao, Junjie Ye, Yuxi Qian, Marco Pavone, and Yue Wang. A Language Agent for Autonomous Driving. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=UPE6WYE8vg.

Hao Sha, Yao Mu, Yuxuan Jiang, Li Chen, Chenfeng Xu, Ping Luo, Shengbo Eben Li, Masayoshi Tomizuka, Wei Zhan, and Mingyu Ding. LanguageMPC: Large Language Models as Decision Makers for Autonomous Driving. *arXiv preprint arXiv:2310.03026*, 2023. doi: 10.48550/arXiv.2310.03026. URL https://arxiv.org/abs/2310.03026.

Shai Shalev-Shwartz, Shaked Shammah, and Amnon Shashua. On a Formal Model of Safe and Scalable Self-Driving Cars. *arXiv preprint arXiv:1708.06374*, 2017. doi: 10.48550/arXiv.1708.06374. URL https://arxiv.org/abs/1708.06374.

Hao Shao, Yuxuan Hu, Letian Wang, Guanglu Song, Steven L Waslander, Yu Liu, and Hongsheng Li. LMDrive: Closed-Loop End-to-End Driving with Large Language Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15120–15130, 2024. doi: 10.1109/CVPR52733.2024.01432.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language Agents with Verbal Reinforcement Learning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 8634–8652, 2023. doi: 10.52202/075280-0377.

Tingting Sui and Sebastian Sunday Grève. A Confucian Algorithm for Autonomous Vehicles. *Science and Engineering Ethics*, 30(6):52, 2024. doi: 10.1007/s11948-024-00514-z.

Tao Tang, Dafeng Wei, Zhengyu Jia, Tian Gao, Changwei Cai, Chengkai Hou, Peng Jia, Kun Zhan, Haiyang Sun, Jingchen Fan, Yixing Zhao, Xiaodan Liang, Xianpeng Lang, and Yang Wang. BEV-TSR: Text-Scene Retrieval in BEV Space for Autonomous Driving. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(7):7275–7283, 2025. doi: 10.1609/aaai.v39i7.32782.

Thomas Tian, Boyi Li, Xinshuo Weng, Yuxiao Chen, Edward Schmerling, Yue Wang, Boris Ivanovic, and Marco Pavone. Tokenize the World into Object-Level Knowledge to Address Long-Tail Events in Autonomous Driving. In *Proceedings of the 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pp. 3656–3673. PMLR, 2025a. URL https://proceedings.mlr.press/v270/tian25b.html.

Yafu Tian, Alexander Carballo, Ruifeng Li, Simon Thompson, and Kazuya Takeda. Query by Example: Semantic Traffic Scene Retrieval Using LLM-Based Scene Graph Representation. *Sensors*, 25(8):2546, 2025b. doi: 10.3390/s25082546.

Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. Voyager: An Open-Ended Embodied Agent with Large Language Models. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=ehfRiF0R3a.

Licheng Wen, Daocheng Fu, Xin Li, Xinyu Cai, Tao Ma, Pinlong Cai, Min Dou, Botian Shi, Liang He, and Yu Qiao. DiLu: A Knowledge-Driven Approach to Autonomous Driving with Large Language Models. In *12th International Conference on Learning Representations, ICLR 2024*, 2024. URL https://openreview.net/forum?id=OqTMUPuLuC.

Zhongyu Xia, Wenhao Chen, Yongtao Wang, and Ming-Hsuan Yang. KnowVal: A Knowledge-Augmented and Value-Guided Autonomous Driving System. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3740–3749, 2026.

Xin Xu, Xinya Lu, Yusi Yan, Mao Tian, and Jianan Wang. ConfuciusLM: Exploring Confucian Personalized Instruction for Primary Learners With Large Language Models. *IEEE Transactions on Learning Technologies*, 19:433–446, 2026. doi: 10.1109/TLT.2026.3688630.

Zhenhua Xu, Yujia Zhang, Enze Xie, Zhen Zhao, Yong Guo, Kwan-Yee K Wong, Zhenguo Li, and Hengshuang Zhao. DriveGPT4: Interpretable End-to-End Autonomous Driving via Large Language Model. *IEEE Robotics and Automation Letters*, 9(10):8186–8193, 2024. doi: 10.1109/LRA.2024.3440097.

Zhenhua Xu, Yan Bai, Yujia Zhang, Zhuoling Li, Fei Xia, Kwan-Yee K Wong, Jianqiang Wang, and Hengshuang Zhao. DriveGPT4-V2: Harnessing Large Language Model Capabilities for Enhanced Closed-Loop Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17261–17270, 2025. doi: 10.1109/CVPR52734.2025.01609.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. ReAct: Synergizing Reasoning and Acting in Language Models. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=WE_vluYUL-X.

Hao Ye, Mengshi Qi, Zhaohong Liu, Liang Liu, and Huadong Ma. SafeDriveRAG: Towards Safe Autonomous Driving with Knowledge Graph-Based Retrieval-Augmented Generation. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 11170–11178, 2025. doi: 10.1145/3746027.3755868.

Linhao Yu, Yongqi Leng, Yufei Huang, Shang Wu, Haixin Liu, Xinmeng Ji, Jiahui Zhao, Jinwang Song, Tingting Cui, Xiaoqing Cheng, Tao Liu, and Deyi Xiong. CMoralEval: A Moral Evaluation Benchmark for Chinese Large Language Models. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 11817–11837, 2024. doi: 10.18653/v1/2024.findings-acl.703.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, Fei Huang, and Jingren Zhou. Qwen3 Embedding: Advancing Text Embedding and Reranking Through Foundation Models. *arXiv preprint arXiv:2506.05176*, 2025. doi: 10.48550/arXiv.2506.05176. URL https://arxiv.org/abs/2506.05176.

Liang Zhao, Bokai Chen, Weiwei Zheng, Lingyan Zhou, and Xiaojun Ding. CyberConfucius: An Interactive Confucian Philosophical Counseling System for Mental Health. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 343–347, 2025. doi: 10.1145/3714394.3754364.

Rui Zhao, Haofeng Hu, Zhenhai Gao, Jiaqiao Liu, and Fei Gao. VLADriver-RAG: Retrieval-Augmented Vision-Language-Action Models for Autonomous Driving. *arXiv preprint arXiv:2605.08133*, 2026. doi: 10.48550/arXiv.2605.08133. URL https://arxiv.org/abs/2605.08133.

Zeyu Zhu and Huijing Zhao. A Survey of Deep RL and IL for Autonomous Driving Policy Learning. *IEEE Transactions on Intelligent Transportation Systems*, 23(9):14043–14065, 2022. doi: 10.1109/TITS.2021.3134702.

A Confucian Classical Works

This appendix lists the 131 Confucian and related classical works used to derive the driving-relevant value principles. Each entry provides the title in Hanyu Pinyin followed by its original Chinese title. Where a well-established English title exists, it is also included; otherwise, the English rendering is a concise descriptive gloss and does not imply the availability of a complete English translation.

1. *Zhouyi* (周易; *Book of Changes*)
2. *Shangshu* (尚书; *Book of Documents*)
3. *Lunyu* (论语; *The Analects*)
4. *Mengzi* (孟子; *Mencius*)
5. *Xunzi* (荀子)
6. *Liji* (礼记; *Book of Rites*)
7. *Daxue* (大学; *The Great Learning*)
8. *Zhongyong* (中庸; *The Doctrine of the Mean*)
9. *Liyun* (礼运; *The Conveyance of Rites*)
10. *Xiaojing* (孝经; *Classic of Filial Piety*)
11. *Xinyu* (新语; *New Discourses*)
12. *Xinshu* (新书; *New Writings*)
13. *Chunqiu fanlu* (春秋繁露; *Luxuriant Dew of the Spring and Autumn Annals*)
14. *Yantielun* (盐铁论; *Discourses on Salt and Iron*)
15. *Shuoyuan* (说苑; *Garden of Persuasions*)
16. *Taixuan* (太玄; *Canon of Supreme Mystery*)
17. *Fayan* (法言; *Exemplary Sayings*)
18. *Xinlun* (新论; *New Treatise*)
19. *Baihu tongyi* (白虎通义; *Comprehensive Discussions in the White Tiger Hall*)
20. *Lunheng* (论衡; *Balanced Discourses*)
21. *Qianfulun* (潜夫论; *Discourses of a Recluse*)
22. *Zhenglun* (政论; *Political Treatise*)
23. *Shenjian* (申鉴; *Extended Reflections*)
24. *Zhonglun* (中论; *Discourses on the Mean*)
25. *Changyan* (昌言; *Broad Discourses*)
26. *Lunyu jijie* (论语集解; *Collected Explanations of the Analects*)
27. *Zhouyi lueli* (周易略例; *General Remarks on the Book of Changes*)
28. *Kongzi jiayu* (孔子家语; *Family Sayings of Confucius*)
29. *Yanshi jiaxun* (颜氏家训; *Family Instructions for the Yan Clan*)
30. *Shitong* (史通; *Generality of Historiography*)
31. *Liu Zongyuan ji* (柳宗元集; *Collected Works of Liu Zongyuan*)
32. *Tianshuo* (天说; *Discourse on Heaven*)
33. *Tiandui* (天对; *Answers to Heaven*)
34. *Fengjian lun* (封建论; *Discourse on Feudalism*)
35. *Han Changli ji* (韩昌黎集; *Collected Works of Han Yu*)
36. *Yuandao* (原道; *Inquiry into the Way*)
37. *Fuxing shu* (复性书; *Book on Returning to One's Nature*)
38. *Liu Yuxi ji* (刘禹锡集; *Collected Works of Liu Yuxi*)
39. *Culai ji* (徂徠集; *Collected Works of Culai*)
40. *Li Gou ji* (李覯集; *Collected Works of Li Gou*)
41. *Chunqiu zunwang fawei* (春秋尊王发微; *Explication of Revering the King in the Spring and Autumn Annals*)
42. *Zhouzi quanshu* (周子全书; *Complete Works of Zhou Dunyi*)
43. *Zhang Zai ji* (张载集; *Collected Works of Zhang Zai*)
44. *Hongfan zhuan* (洪范传; *Commentary on the Great Plan*)
45. *Er Cheng ji* (二程集; *Collected Works of the Cheng Brothers*)

46. *Qianxu* (潜虚; *Hidden Void*)
47. *Dongpo Yizhuan* (东坡易传; *Dongpo's Commentary on the Book of Changes*)
48. *Guishan ji* (龟山集; *Collected Works of Guishan*)
49. *Zhiyan* (知言; *Known Words*)
50. *Hengpu ji* (横浦集; *Collected Works of Hengpu*)
51. *Nanxuan ji* (南轩集; *Collected Works of Nanxuan*)
52. *Donglai ji* (东莱集; *Collected Works of Donglai*)
53. *Xingli zixun* (性理字训; *Glossary of Nature and Principle*)
54. *Lu Jiuyuan ji* (陆九渊集; *Collected Works of Lu Jiuyuan*)
55. *Chen Liang ji* (陈亮集; *Collected Works of Chen Liang*)
56. *Zhuzi yulei* (朱子语类; *Classified Conversations of Master Zhu*)
57. *Yiluo yuanyuan lu* (伊洛渊源录; *Record of the Origins of Yiluo Learning*)
58. *Sishu jizhu* (四书集注; *Collected Commentaries on the Four Books*)
59. *Hui'an ji* (晦庵集; *Collected Works of Hui'an*)
60. *Jinsi lu* (近思录; *Reflections on Things at Hand*)
61. *Chengzhai Yizhuan* (诚斋易传; *Chengzhai's Commentary on the Book of Changes*)
62. *Ye Shi ji* (叶适集; *Collected Works of Ye Shi*)
63. *Xixue jiyan* (习学记言; *Record of Words from Sustained Study*)
64. *Beixi ziyi* (北溪字义; *Neo-Confucian Terms Explained*)
65. *Cihu yishu* (慈湖遗书; *Posthumous Writings of Cihu*)
66. *Zhen Xishan quanji* (真西山全集; *Complete Works of Zhen Xishan*)
67. *Daxue yanyi* (大学衍义; *Expanded Meaning of the Great Learning*)
68. *Heshan quanji* (鹤山全集; *Complete Works of Heshan*)
69. *Huangshi richao* (黄氏日抄; *Daily Notes of Master Huang*)
70. *Wenshan xiansheng quanji* (文山先生全集; *Complete Works of Master Wenshan*)
71. *Kunxue jiwen* (困学纪闻; *Record of Knowledge Acquired through Hard Study*)
72. *Luzhai yishu* (鲁斋遗书; *Posthumous Writings of Luzhai*)
73. *Liu Jingxiu xiansheng wenji* (刘静修先生文集; *Collected Works of Master Liu Jingxiu*)
74. *Boya qin* (伯牙琴; *Boya's Zither*)
75. *Caolu Wu Wenzhenggong wenji* (草庐吴文正公文集; *Collected Works of Wu Wenzheng of Caolu*)
76. *Yulizi* (郁离子; *Master Yuli*)
77. *Lixue leibian* (理学类编; *Classified Compilation of the Learning of Principle*)
78. *Xingli daquanshu* (性理大全书; *Great Compendium of Nature and Principle*)
79. *Xue Jingxuan xiansheng wenji* (薛敬轩先生文集; *Collected Works of Master Xue Jingxuan*)
80. *Dushu lu* (读书录; *Record of Reading*)
81. *Kangzhai wenji* (康斋文集; *Collected Works of Kangzhai*)
82. *Juye lu* (居业录; *Record of Dwelling in One's Vocation*)
83. *Baishazi quanji* (白沙子全集; *Complete Works of Master Baisha*)
84. *Yangming quanshu* (阳明全书; *Complete Works of Wang Yangming*)
85. *Chuanxilu* (传习录; *Instructions for Practical Living*)
86. *Wang Xinzhai xiansheng yiji* (王心斋先生遗集; *Posthumous Works of Master Wang Xinzhai*)
87. *Jingyezi neipian* (泾野子内篇; *Inner Chapters of Master Jingye*)
88. *Bozhai ji* (柏斋集; *Collected Works of Bozhai*)
89. *Shenyan* (慎言; *Cautious Words*)
90. *Yashu* (雅述; *Elegant Expositions*)
91. *Kunzhi ji* (困知记; *Record of Knowledge Attained through Difficulty*)
92. *Zheng'an cungao* (整庵存稿; *Surviving Drafts of Zheng'an*)
93. *Mingdao bian* (明道编; *Compilation on Clarifying the Way*)
94. *Ganquan wenji* (甘泉文集; *Collected Works of Ganquan*)
95. *Gewu tong* (格物通; *Comprehensive Study of the Investigation of Things*)
96. *Kunbian lu* (困辨录; *Record of Deliberations amid Difficulty*)
97. *Wenbian lu* (问辨录; *Record of Questions and Distinctions*)
98. *He Xinyin ji* (何心隐集; *Collected Works of He Xinyin*)
99. *Zhang Taiyue quanji* (张太岳全集; *Complete Works of Zhang Taiyue*)
100. *Xiaoxinzai zhaji* (小心斋箚记; *Notes from the Xiaoxin Studio*)
101. *Shenyin yu* (呻吟语; *Groanings and Mutterings*)
102. *Liuzi quanji* (刘子全集; *Complete Works of Master Liu*)
103. *Rongtan wenye* (榕坛问业; *Questions on Learning at Rongtan*)
104. *Wuhe ji* (无何集; *Wuhe Collection*)
105. *Lixue zongzhuan* (理学宗传; *Genealogical Transmission of the Learning of Principle*)
106. *Xiafeng xiansheng ji* (夏峰先生集; *Collected Works of Master Xiafeng*)
107. *Chen Que ji* (陈确集; *Collected Works of Chen Que*)
108. *Daxue bian* (大学辨; *Discussions of the Great Learning*)
109. *Guyan* (瞽言; *Blind Words*)
110. *Zhu Shunshui ji* (朱舜水集; *Collected Works of Zhu Shunshui*)

111. *Rizhi lu* (日知录; *Record of Daily Knowledge*)
112. *Gu Tinglin shiwen ji* (顾亭林诗文集; *Collected Poetry and Prose of Gu Tinglin*)
113. *Zhangzi Zhengmeng zhu* (张子正蒙注; *Commentary on Master Zhang's Correcting Youthful Ignorance*)
114. *Shangshu yinyi* (尚书引义; *Notes on the Meanings of the Book of Documents*)
115. *Zhouyi waizhuan* (周易外传; *Outer Commentary on the Book of Changes*)
116. *Du Tongjian lun* (读通鉴论; *Reflections on Reading the Comprehensive Mirror*)
117. *Mingyi daifang lu* (明夷待访录; *Waiting for the Dawn: A Plan for the Prince*)
118. *Mingru xue'an* (明儒学案; *Records of Ming Scholars*)
119. *Song-Yuan xue'an* (宋元学案; *Records of Song and Yuan Scholars*)
120. *Hongdao shu* (弘道书; *Book on Expanding the Way*)
121. *Qianshu* (潜书; *Hidden Writings*)
122. *Sicun bian* (四存编; *Four Works on Preservation*)
123. *Sishu zhengwu* (四书正误; *Corrections of Errors in the Four Books*)
124. *Erqu ji* (二曲集; *Collected Works of Erqu*)
125. *Daxue bianye* (大学辨业; *Discriminating the Meaning of the Great Learning*)
126. *Shugu houji* (恕谷后集; *Later Collection of Shugu*)
127. *Yuanshan* (原善; *Inquiry into Goodness*)
128. *Mengzi ziyi shuzheng* (孟子字义疏证; *Evidential Study of the Meanings of Terms in Mencius*)
129. *Yiyan* (意言; *Words of Intent*)
130. *Diaogu ji* (雕菰集; *Collected Writings from the Diaogu Studio*)
131. *Hanxue shangdui* (汉学商兑; *A Critical Assessment of Han Learning*)

B Prompt Templates and Closed-Loop Decision Example

This appendix presents the prompt templates used in the main and ablation experiments, together with a representative closed-loop decision example. As shown in Figure 4, the model input consists of a system prompt and a user prompt. The system prompt specifies the model's role as a driving assistant, the requirements for value-guided reasoning, and the output format, while the user prompt provides the current driving scenario, the three historical cases retrieved by PSSR, and the currently available driving actions. To further illustrate how this prompt is instantiated at a decision step, Figure 5 presents a current driving scenario and one representative retrieved historical case, and Figure 6 shows the corresponding LLM decision response.

System Prompt:

You are a professional driving assistant well-versed in Confucian wisdom.
Your task is to read the current driving scenario, refer to the provided historically similar cases, and make the most reasonable decision.
You will be given a detailed description of the current-frame driving scenario, three retrieved historical cases, and a list of available driving action IDs.
When making the decision, you should comprehensively assess the current driving scenario and deeply refer to the three retrieved historical cases. You need to discern how the driving-relevant value principles in the retrieved cases guide the reasoning process, identify the value principles most relevant to the current road condition, and use them to guide the subsequent reasoning process.
Then, naturally generate a reasoning process that integrates philosophical wisdom with the physical driving context.
After clearly deriving the action to be executed, you must strictly end your response with `#### <action ID>`.
Please strictly obey traffic rules and prioritize absolute safety.

User Prompt:

The information for the current scenario is as follows:
{scenario description}
The information for the three retrieved historical cases is as follows:
{retrieved historical cases}
The available actions are as follows:
{available actions}
Please refer to these cases flexibly and deliberate carefully to make your final decision.

Figure 4: System and user prompts used by CPW-Drive for closed-loop driving decision-making.

Scenario Description:

You are driving on the road with a speed of 21.68 m/s and an acceleration of -1.28 m/s^2. Vehicle A is ahead of you in the same lane, 42.73m away, with a speed of 20.83 m/s and an acceleration of -1.15 m/s^2. Vehicle B is ahead of you in the left lane, 4.20m away, with a speed of 18.84 m/s and an acceleration of -2.58 m/s^2.

Historical Reference Scenarios:

Scenario: You are driving on the road with a speed of 21.12 m/s and an acceleration of -0.22 m/s^2. Vehicle A is ahead of you in the same lane, 42.75 m away, with a speed of 19.07 m/s and an acceleration of -3.47 m/s^2. Vehicle B is ahead of you in the left lane, 9.85 m away, with a speed of 20.71 m/s and an acceleration of -0.89 m/s^2. Vehicle C is ahead of you in the right lane, 32.62 m away, with a speed of 18.25 m/s and an acceleration of -1.92 m/s^2.
Value Principles: Wisdom, Propriety, Moderation, Empathy, Respect, Righteousness
Reasoning Process: On the physical level, although Vehicle A ahead is decelerating relatively quickly at -3.47 m/s^2, the current longitudinal distance of 42.75 m still provides sufficient reaction space and does not require an immediate evasive maneuver. From the perspective of Wisdom, the current lane retains an adequate safety buffer. Given that Vehicle B on the left and Vehicle C on the right are both ahead, maintaining the current lane reflects Propriety by preserving stable traffic flow. Moderation supports avoiding an excessive response to the leading vehicle's temporary deceleration and maintaining a balanced driving state. This action also reflects Empathy toward the leading vehicle's braking behavior, Respect for traffic order, and Righteousness in balancing ego-vehicle safety with the interests of other road users.
Driving Action: Keep unchanged

Figure 5: Example of a current driving scenario and one representative historical case retrieved by PSSR.

LLM Decision Response:

Guided by Confucian philosophy, I have analyzed the current scenario in conjunction with similar historical cases. My current speed is 21.68 m/s. Vehicle A, ahead in the same lane, is 42.73 m away; although it is decelerating slightly, the gap remains ample and poses no imminent threat. Meanwhile, Vehicle B ahead in the left lane is a mere 4.20 m away and is also decelerating. A rash left lane change would instantly trigger a collision due to the extremely short distance and speed differential, conflicting with the judgment guided by Wisdom and the order reflected in Propriety. Furthermore, although the right lane is clear, a lane change is unnecessary because the current lane remains safe, and the principle of Moderation favors maintaining stability over making an unnecessary maneuver. By applying Empathy to accommodate the leading vehicle and Righteousness to protect the overall right-of-way, maintaining the current lane is sufficient to sustain a dynamic equilibrium. Therefore, I should adopt the strategy of remaining unchanged, avoiding the escalation of risks, and reflecting Benevolence while prioritizing safety.
1

Figure 6: LLM decision response generated from the current driving scenario and the retrieved historical cases. The response provides the reasoning process and ends with the action ID in the required output format.

Figures 7 and 8 present the system prompts used for the **Philosophy in Prompt** and **Conservative Prompt** settings in the ablation study, respectively.

System Prompt:

You are a professional driving assistant well-versed in Confucian wisdom.
Your task is to read the current driving scenario and make the most reasonable decision.
You will be given a detailed description of the current-frame driving scenario and a list of available driving action IDs.
When making the decision, you should comprehensively assess the current driving scenario and use the following driving-relevant value principles as high-level guidance: Wisdom, Righteousness, Respect, Adaptability, Benevolence, Courage, Moderation, Propriety, Empathy, and Trustworthiness.
Identify the value principles most relevant to the current road condition and use them to guide the reasoning process. Then, naturally generate a reasoning process that integrates philosophical wisdom with the physical driving context.
After clearly deriving the action to be executed, you must strictly end your response with `#### <action ID>`.
Please strictly obey traffic rules and prioritize absolute safety.

Figure 7: System prompt used for the **Philosophy in Prompt** setting.

System Prompt:

You are a professional driving assistant.
Your task is to read the current driving scenario, refer to the provided historically similar cases, and make the most reasonable decision.
You will be given a detailed description of the current-frame driving scenario, three retrieved historical cases, and a list of available driving action IDs.
When making the decision, you should comprehensively assess the current driving scenario and refer to the three retrieved historical cases and their reasoning processes.
Adopt a conservative driving strategy: when not necessary, avoid lane changes, acceleration, and other proactive maneuvers, and prefer maintaining the current driving state when it is safe to do so.
After clearly deriving the action to be executed, you must strictly end your response with `#### <action ID>`.
Please strictly obey traffic rules and prioritize absolute safety.

Figure 8: System prompt used for the **Conservative Prompt** setting.

C External Evaluation of Confucian Value Conformity

This appendix further evaluates the extent to which the reasoning processes and driving decisions generated by CPW-Drive conform to Confucian values. To enhance the independence and objectivity of the evaluation, we adopt the publicly available evaluation framework from PICACO as a third-party assessment tool rather than devising evaluation criteria specifically for our method. PICACO, published at ICML 2026, investigates pluralistic value alignment and explicitly treats Confucianism as a value composition rooted in a specific cultural and philosophical tradition (Jiang et al., 2026). It represents Confucian values along four dimensions, namely Benevolence, Conformity, Tradition, and Security, thereby providing an evaluation module that measures the extent to which a model response reflects these values. We do not use PICACO's meta-instruction optimization method and employ only its Confucian-value evaluation module. The entire evaluation directly runs the official publicly available code, without any modification to the evaluation prompts, value definitions, scoring criteria, or aggregation procedure.

Specifically, we randomly sampled one decision frame from each of 40 completed runs, each initialized with a random map seed, yielding 40 evaluation samples. Because adjacent frames within the same driving episode typically contain highly similar traffic scenes and decision contexts, only one frame was sampled from each run to reduce temporal redundancy and prevent nearly identical scenes from being overrepresented in the evaluation set. For each sample, the original driving decision query sent to CPW-Drive contained the current traffic scene information, and the complete reasoning process and final driving decision generated by CPW-Drive constituted the model response to be evaluated. The Confucian value evaluation module in PICACO was provided with both the original query and the model response and assessed the extent to which the response reflected Confucian values in the context of the scene described in the query.

The PICACO evaluation module scores each model response along four dimensions, namely Benevolence, Conformity, Tradition, and Security, and also provides a relevance score that measures the correspondence between the model response and the original query. All scores use a scale from 1 to 5, where a higher score indicates that the response more strongly reflects the corresponding Confucian value or is more relevant to the original query. The module first averages the scores across the four Confucian value dimensions, then multiplies this average by the relevance score and divides the result by the maximum score of 5 to obtain the overall Confucian conformity score. The final score therefore jointly reflects the extent to which the response embodies Confucian values and its relevance to the original query. The distributions of these scores across the 40 evaluation samples are shown in Figure 9.

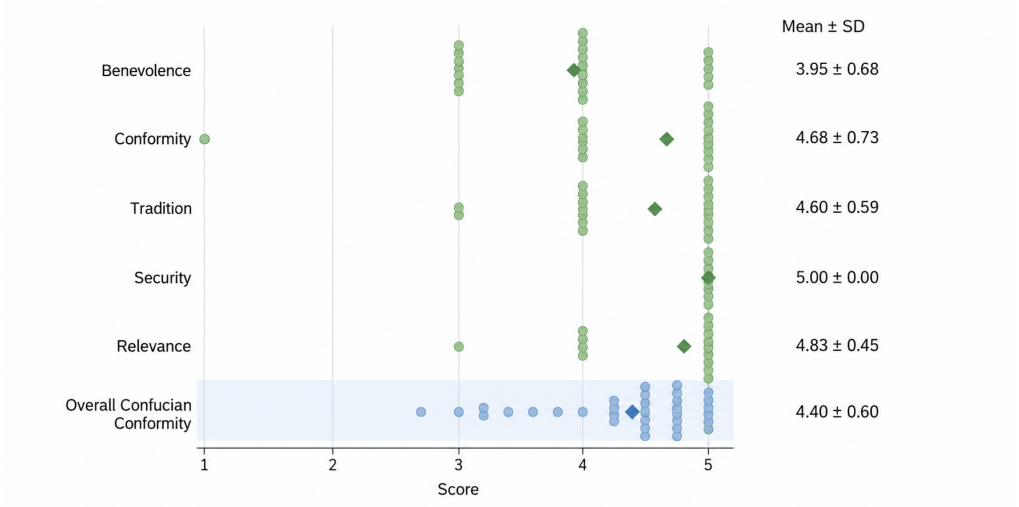

Figure 9: Distribution of PICACO-based Confucian value evaluation scores for CPW-Drive. Each circle represents one of the 40 sampled decision frames, diamonds denote the mean scores, and the values on the right report the mean ± standard deviation. The shaded row shows the overall Confucian conformity score obtained by combining the four value dimensions with response relevance.

Across the 40 evaluation samples, CPW-Drive achieved scores of 3.95 ± 0.68, 4.68 ± 0.73, 4.60 ± 0.59, and 5.00 ± 0.00 for Benevolence, Conformity, Tradition, and Security, respectively. The relevance score was 4.83 ± 0.45, and the overall Confucian conformity score was 4.40 ± 0.60. Security received the maximum score for every sample, while Conformity and Tradition also remained at high levels. The high relevance score indicates that the model responses were generally consistent with the original queries. Overall, the third-party evaluation indicates that the reasoning processes and driving decisions generated by CPW-Drive exhibit a high degree of conformity with Confucian values.

D Qualitative Retrieval Case Analysis

This appendix provides a qualitative comparison between text-vector retrieval and Physics-aware Spatial Similarity Retrieval (PSSR), using the representative highway interaction scenario shown in Figure 10(a) as the query. The scenario contains a same-lane front vehicle, a left-lane front vehicle, and a right-lane rear vehicle, capturing longitudinal car-following constraints and adjacent-lane interactions. For the same query, we apply text-vector retrieval based on Qwen3-Embedding-8B and PSSR, and show the three highest-ranked historical cases returned by each method in Figure 10(b)–(g).

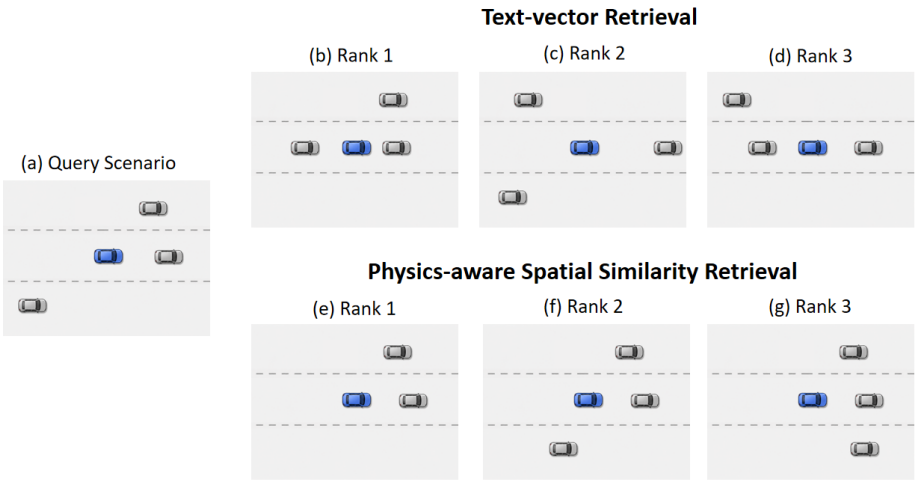

Figure 10: Qualitative comparison of historical scene retrieval results. (a) Query scenario. (b)–(d) The three highest-ranked cases retrieved by text-vector retrieval. (e)–(g) The three highest-ranked cases retrieved by Physics-aware Spatial Similarity Retrieval.

Figure 10(b)–(d) shows that text-vector retrieval returns historical cases with the same number of surrounding vehicles as the query, but does not consistently preserve their longitudinal positions or lane relations. Two factors may explain this behavior. First, all scenarios are rendered using highly similar fixed text templates, with each surrounding vehicle contributing a similarly structured clause. Scenes containing the same number of surrounding vehicles therefore have similar text lengths and sentence structures, and may receive high overall text similarity even when their physical configurations differ. Second, text-vector retrieval does not explicitly model ego-centered vehicle geometry, and its similarity measure is not directly constrained by lane relations, relative distances, or motion states. Consequently, spatial differences that occupy only a small portion of the description but are critical to driving decisions may be weakened by the overall semantic similarity.

In contrast, the PSSR results in Figure 10(e)–(g) are determined directly from ego-centered vehicle physical states and do not depend on the length or template structure of the scene descriptions. Although these historical cases differ from the query in vehicle count and exact positions, their principal vehicle relations and lane constraints are better preserved. PSSR neither requires two scenes to contain identical vehicle sets nor establishes explicit one-to-one vehicle correspondences; instead, it measures scene similarity according to the consistency of the overall spatial configurations around the ego vehicle. By combining the current and velocity-extrapolated spaces, PSSR accounts for both the current vehicle layout and short-term motion trends, thereby providing the LLM with historical decision references under more closely matched physical constraints.

Overall, this qualitative example illustrates that the two retrieval methods measure scene similarity at different levels: text-vector retrieval primarily relies on the overall linguistic semantics of scene descriptions, whereas PSSR explicitly compares the current and velocity-extrapolated vehicle configurations.

E Computational Complexity and Runtime Efficiency of CPW-Drive

This appendix analyzes the computational complexity and working-memory cost of PSSR and empirically compares its retrieval efficiency with vector-based retrieval methods, complementing the method description in Section 3.4.

Let the historical driving knowledge base contain M cases, with each scene containing at most N surrounding vehicles. For a current scene and a single historical scene, the scene-level spatial overlap in PSSR requires pairwise spatial similarity computation between the two vehicle sets, resulting in a computational complexity of $O(N^2)$. Computing both the current and velocity-extrapolated spatial

Table 3: Online decision latency comparison of different methods. LLM inference time and total time are reported as mean ± standard deviation; P95 denotes the 95th percentile of the total time, and ↓ indicates lower is better.

Method	LLM Inference Time (s)↓	Total Time (s)↓	P95 Total Time (s)↓
DriveLikeAHuman	$\mathbf{5.24 \pm 2.23}$	$\mathbf{5.24 \pm 2.23}$	$\mathbf{9.77}$
DiLu	13.55 ± 6.06	14.81 ± 6.18	24.52
LimSim++	39.79 ± 24.43	42.30 ± 25.28	89.99
CPW-Drive (Ours)	20.06 ± 10.05	20.06 ± 10.05	40.71

Table 4: Historical scene retrieval time comparison of different methods. Retrieval time is reported as mean ± standard deviation, and ↓ indicates lower is better.

Metric	DiLu	LimSim++	Local Vector Retrieval	**CPW-Drive (Ours)**
Retrieval Time (ms)↓	1257.28 ± 1145.53	1506.31 ± 1867.74	20.96 ± 2.64	$\mathbf{4.23 \pm 1.05}$

states introduces only a fixed number of additional operations and therefore does not change the asymptotic complexity. Comparing the current scene with all M historical scenes thus requires $O(MN^2)$ time. After all similarity scores are computed, the M scores are fully sorted and the Top-k historical cases are selected. The sorting step requires $O(M \log M)$ time, while selecting the first k results incurs an additional $O(k)$ cost. Since this term is small relative to the preceding similarity computation and sorting costs, the overall online retrieval complexity of PSSR can be written as $O(MN^2 + M \log M)$. In terms of working memory, the retrieval process stores the vehicle states of the current scene together with the similarity scores for all M historical cases, resulting in an additional working-memory complexity of $O(M + N)$.

We next evaluate the online decision latency of all methods to quantify the computational overhead introduced by different inference and retrieval pipelines. To control for differences in the LLM backend, all methods use the same DeepSeek-V4-Pro API to generate driving decisions while retaining their original prompts and retrieval configurations. DriveLikeAHuman does not include an explicit historical scene retrieval module. Following their open-source implementations, DiLu and LimSim++ both use text-embedding-ada-002 for semantic retrieval and store 20 historical examples in their retrieval libraries. In contrast, CPW-Drive does not rely on an additional embedding model; instead, it directly retrieves relevant scenes from a knowledge base of 135 verified historical examples using PSSR. We report the LLM inference time, the end-to-end total decision time, and the 95th percentile (P95) of the total decision latency. The average latency metrics are reported as mean ± standard deviation, while P95 characterizes the long-tail latency of the complete decision pipeline.

Table 3 shows that DriveLikeAHuman has the lowest mean inference latency, followed by DiLu, CPW-Drive, and LimSim++. These measurements compare complete method configurations rather than isolated model throughput because their prompts and reasoning pipelines differ. CPW-Drive requires 20.06 ± 10.05 s per decision on average, which exceeds the 1 Hz simulated policy interval. The large standard deviations also show substantial wall-clock variability from scene, network, or server-side factors. The current implementation should therefore be interpreted as a simulation prototype rather than a real-time controller. We further compare the retrieval modules themselves.

Because DiLu and LimSim++ call a remote embedding API and use 20-case retrieval libraries in their original implementations, their reported retrieval times reflect the realized latency of the original retrieval pipelines, including network and service variability, rather than a controlled comparison of retrieval algorithms. For a more controlled local comparison, we deploy Qwen3-Embedding-8B (Zhang et al., 2025) and search the same 135 cases used by PSSR. Table 4 shows that local vector retrieval is numerically faster than the two remote pipelines but slower than PSSR. PSSR directly computes physical similarity without vectorization and has the lowest mean retrieval latency at 4.23 ± 1.05 ms. Relative to the 20.06 s end-to-end mean, retrieval contributes only a small fraction of CPW-Drive's wall-clock latency.

F Derivation of the Spatial Kernel and Normalized Similarity

This appendix provides detailed derivations of the main mathematical formulations underlying PSSR, complementing the method description in Section 3.4. These derivations cover the Gaussian-overlap spatial kernel, the cancellation of the common multiplicative constant, the determination of the lateral spatial scale, and the normalized similarity and its value range.

In the main text, the position of each surrounding vehicle is represented by a two-dimensional anisotropic Gaussian density,

$$\rho_i(x,y) = \exp\left(-\frac{(x-x_i)^2}{2\sigma_x^2} - \frac{(y-y_i)^2}{2\sigma_y^2}\right).$$

For vehicles i and j, the spatial overlap between their Gaussian densities is defined as

$$I_{ij} = \iint \rho_i(x,y)\rho_j(x,y)\,dxdy.$$

Because the exponential term is separable along the longitudinal and lateral coordinates, the two-dimensional integral can be written as the product of two one-dimensional integrals:

$$I_{ij} = I_x(i,j)I_y(i,j).$$

The longitudinal integral is

$$I_x(i,j) = \int_{-\infty}^{\infty} \exp\left(-\frac{(x-x_i)^2}{2\sigma_x^2} - \frac{(x-x_j)^2}{2\sigma_x^2}\right) dx.$$

Let

$$m_x = \frac{x_i + x_j}{2}.$$

Then

$$(x-x_i)^2 + (x-x_j)^2 = 2(x-m_x)^2 + \frac{(x_i-x_j)^2}{2}.$$

Therefore,

$$\begin{aligned} I_x(i,j) &= \exp\left(-\frac{(x_i-x_j)^2}{4\sigma_x^2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{(x-m_x)^2}{\sigma_x^2}\right) dx \\ &= \sqrt{\pi}\sigma_x \exp\left(-\frac{(x_i-x_j)^2}{4\sigma_x^2}\right). \end{aligned}$$

Similarly, the lateral integral is

$$I_y(i,j) = \sqrt{\pi}\sigma_y \exp\left(-\frac{(y_i-y_j)^2}{4\sigma_y^2}\right).$$

Thus, the complete two-dimensional Gaussian-overlap integral is

$$I_{ij} = \pi\sigma_x\sigma_y \exp\left(-\frac{(x_i-x_j)^2}{4\sigma_x^2} - \frac{(y_i-y_j)^2}{4\sigma_y^2}\right). \tag{8}$$

The spatial kernel used in the main text omits the common multiplicative constant $\pi\sigma_x\sigma_y$ and is defined as

$$K_{\text{space}}(i,j) = \frac{I_{ij}}{\pi\sigma_x\sigma_y} = \exp\left(-\frac{(x_i - x_j)^2}{4\sigma_x^2} - \frac{(y_i - y_j)^2}{4\sigma_y^2}\right).$$

We next show that omitting this common constant does not affect the final normalized similarity. Let the spatial kernel including the full multiplicative constant be

$$\widetilde{K}_{\text{space}}(i,j) = cK_{\text{space}}(i,j), \qquad c = \pi\sigma_x\sigma_y.$$

The corresponding scene-level overlap between scenes A and B is then

$$\widetilde{S}_{\text{space}}(A,B) = \sum_{i=1}^{N_A}\sum_{j=1}^{N_B}\widetilde{K}_{\text{space}}(i,j) = cS_{\text{space}}(A,B).$$

Substituting this expression into the normalized similarity gives

$$\widetilde{K}^0_{\text{space}}(A,B) = \frac{\widetilde{S}_{\text{space}}(A,B)}{\sqrt{\widetilde{S}_{\text{space}}(A,A)\widetilde{S}_{\text{space}}(B,B)}}.$$

Expanding the expression yields

$$\begin{aligned}\widetilde{K}^0_{\text{space}}(A,B) &= \frac{cS_{\text{space}}(A,B)}{\sqrt{cS_{\text{space}}(A,A)\cdot cS_{\text{space}}(B,B)}} \\ &= \frac{S_{\text{space}}(A,B)}{\sqrt{S_{\text{space}}(A,A)S_{\text{space}}(B,B)}} \\ &= K^0_{\text{space}}(A,B).\end{aligned} \tag{9}$$

Therefore, the common multiplicative constant $\pi\sigma_x\sigma_y$ cancels completely during normalization. Omitting this constant changes only the scale of the unnormalized overlap and does not affect the final normalized similarity or the ranking of retrieved historical scenes.

The longitudinal spatial scale σ_x is used to characterize front–rear positional differences between vehicles and is empirically set to $\sigma_x = 5\,\text{m}$. The lateral spatial scale σ_y is determined according to the desired similarity decay over one lane width. Let the lane width be L. When two vehicles differ only by a lateral displacement $\Delta y = L$, we require the spatial kernel to decay to q_{lane} relative to the case $\Delta y = 0$:

$$\frac{K_{\text{space}}(\Delta y = L)}{K_{\text{space}}(\Delta y = 0)} = \exp\left(-\frac{L^2}{4\sigma_y^2}\right) = q_{\text{lane}}.$$

This gives

$$-\frac{L^2}{4\sigma_y^2} = \ln q_{\text{lane}},$$

and therefore

$$\sigma_y = \frac{L}{2\sqrt{-\ln q_{\text{lane}}}}. \tag{10}$$

We set $L = 3.5\,\text{m}$ and $q_{\text{lane}} = 0.10$, giving

$$\sigma_y = \frac{3.5}{2\sqrt{-\ln(0.10)}} \approx 1.1533\,\mathrm{m}.$$

We next show that the scene-level normalization used in the main text can be interpreted as a cosine similarity in feature space and establish the value range of the final PSSR similarity. Following the algorithmic condition specified in the main text, we consider only non-empty scenes containing at least one surrounding vehicle.

Let $c = \pi\sigma_x\sigma_y$, and define the normalized density feature as

$$\phi(p_i) = \frac{\rho_i}{\sqrt{c}}, \qquad p_i = (x_i, y_i).$$

From the Gaussian-overlap derivation above,

$$\langle \phi(p_i), \phi(p_j) \rangle_{L_2} = K_{\mathrm{space}}(i, j).$$

We further define the aggregated scene-level feature representations as

$$\Phi(A) = \sum_{i=1}^{N_A} \phi(p_i^A), \qquad \Phi(B) = \sum_{j=1}^{N_B} \phi(p_j^B).$$

Then

$$\langle \Phi(A), \Phi(B) \rangle = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \langle \phi(p_i^A), \phi(p_j^B) \rangle = S_{\mathrm{space}}(A, B).$$

Similarly,

$$S_{\mathrm{space}}(A, A) = \|\Phi(A)\|^2, \qquad S_{\mathrm{space}}(B, B) = \|\Phi(B)\|^2.$$

Hence, the normalized similarity used in the main text can be written as

$$K^0_{\mathrm{space}}(A, B) = \frac{\langle \Phi(A), \Phi(B) \rangle}{\|\Phi(A)\| \|\Phi(B)\|}. \tag{11}$$

It should be noted that

$$S_{\mathrm{space}}(A, A) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_A} K_{\mathrm{space}}(i, j)$$

contains not only the self-overlap terms of individual vehicles but also cross terms between distinct vehicles within the same scene, and therefore cannot in general be simplified to the number of vehicles N_A.

For non-empty vehicle sets, the normalization denominator is positive. Because the Gaussian spatial kernel is strictly positive for finite vehicle coordinates, the normalized similarity is strictly positive; the Cauchy–Schwarz inequality gives the upper bound:

$$0 < K^0_{\mathrm{space}}(A, B) \leq 1.$$

Therefore, the current-space similarity and velocity-extrapolated-space similarity defined in the main text satisfy

$$K_{\mathrm{cur}}(A,B) \in (0,1], \qquad K_{\mathrm{ext}}(A,B) \in (0,1].$$

Because the two terms are combined using non-negative weights satisfying

$$w_{\mathrm{cur}} + w_{\mathrm{ext}} = 1,$$

the final PSSR similarity is defined as

$$K(A,B) = w_{\mathrm{cur}} K_{\mathrm{cur}}(A,B) + w_{\mathrm{ext}} K_{\mathrm{ext}}(A,B),$$

and therefore

$$K(A,B) \in (0,1]. \tag{12}$$