

Article

Not peer-reviewed version

ConsciousDriver: A Context-Aware Multimodal Personalized Autonomous Driving System

Noah Fang and [Salma Ali](#)*

Posted Date: 20 January 2026

doi: 10.20944/preprints202601.1454.v1

Keywords: autonomous driving; MLLM; context-awareness; personalization; risk adaptation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Article

ConsciousDriver: A Context-Aware Multimodal Personalized Autonomous Driving System

Noah Fang and Salma Ali *

Universiti Teknologi Malaysia

* Correspondence: hva180033@siswa365.um.edu.my

Abstract

Current multimodal large language model (MLLM)-based autonomous driving systems struggle with deep contextual understanding, fine-grained personalization, and transparent risk assessment in complex real-world scenarios. This paper introduces ConsciousDriver, a novel context-aware multimodal personalized autonomous driving system designed to address these limitations. Our system integrates a Context-Awareness Module for richer environmental understanding and a Dynamic Risk Adaptation Mechanism to flexibly modulate driving behaviors based on real-time user prompts and situational risks. Built upon an extended MLLM architecture, ConsciousDriver processes environmental inputs and user prompts to generate deep contextual understanding, context-adaptive danger levels, optimal action decisions, and explicit decision intent explanations. Evaluated on an extended PAD-Highway benchmark, ConsciousDriver demonstrates superior performance in driving safety, efficiency, lane-keeping, and traffic density adaptation. Furthermore, it exhibits robust adaptability to diverse personalized prompts and enhanced performance in challenging traffic scenarios, with lower collision and higher completion rates. Human evaluation confirms the high quality of its explanations. ConsciousDriver represents a significant advancement towards intelligent, adaptive, and trustworthy autonomous driving.

Keywords: autonomous driving; MLLM; context-awareness; personalization; risk adaptation

1. Introduction

Autonomous driving technology has witnessed remarkable advancements, particularly in controlled environments such as highways and well-mapped urban areas. However, extending these capabilities to complex, dynamic, and unpredictable real-world scenarios, which often demand highly personalized and nuanced decision-making, remains a significant challenge [1]. The integration of multimodal large language models (MLLMs) has emerged as a promising paradigm for end-to-end autonomous driving, exemplified by pioneering works like *PADriver* [2]. These systems leverage textual prompts to enable rudimentary personalization of driving styles, such as "slow," "normal," or "fast" modes. While a crucial step forward, current approaches still exhibit several limitations that hinder their deployment in diverse and human-centric driving situations.

The primary challenges in current MLLM-based autonomous driving systems include:

1. **Insufficient Depth of Contextual Understanding:** Existing methods often rely on bird's-eye-view (BEV) frames and simplistic scene descriptions. This limited input prevents a comprehensive grasp of complex environmental dynamics, potential risk sources, and their underlying causal relationships. For instance, distinguishing between a pedestrian merely standing by the roadside and one about to cross requires deeper contextual inference beyond static visual cues.
2. **Limited Personalization Granularity:** The predefined "slow," "normal," or "fast" driving modes offer insufficient flexibility to meet the dynamic and fine-grained preferences of human drivers. A driver's desired style can vary significantly based on time of day, road conditions, traffic density, and specific intentions (e.g., "on this road section, pay attention to pedestrians and slow

down, but try to maintain the flow of traffic"). Existing systems struggle to incorporate such rich, dynamic contextual prompts.

3. **Lack of Transparency and Explainability in Risk Perception and Decision-Making:** While MLLMs can generate danger levels for various actions, the underlying reasoning processes are often opaque. Furthermore, these models do not fully leverage nuanced contextual information for more refined risk assessment and adaptive behavior. This lack of transparency can erode user trust and hinder debugging in critical scenarios.

Motivated by these limitations, our research introduces **ConsciousDriver: A Context-Aware Multi-modal Personalized Autonomous Driving System**. ConsciousDriver aims to elevate personalized autonomous driving to a new level by incorporating a **Context-Awareness Module (CAM)** and a **Dynamic Risk Adaptation Mechanism (DRAM)**. This allows the system to intelligently understand intricate driving intentions, rigorously assess situational risks, and make decisions that are not only safer and more comfortable but also deeply aligned with user expectations.

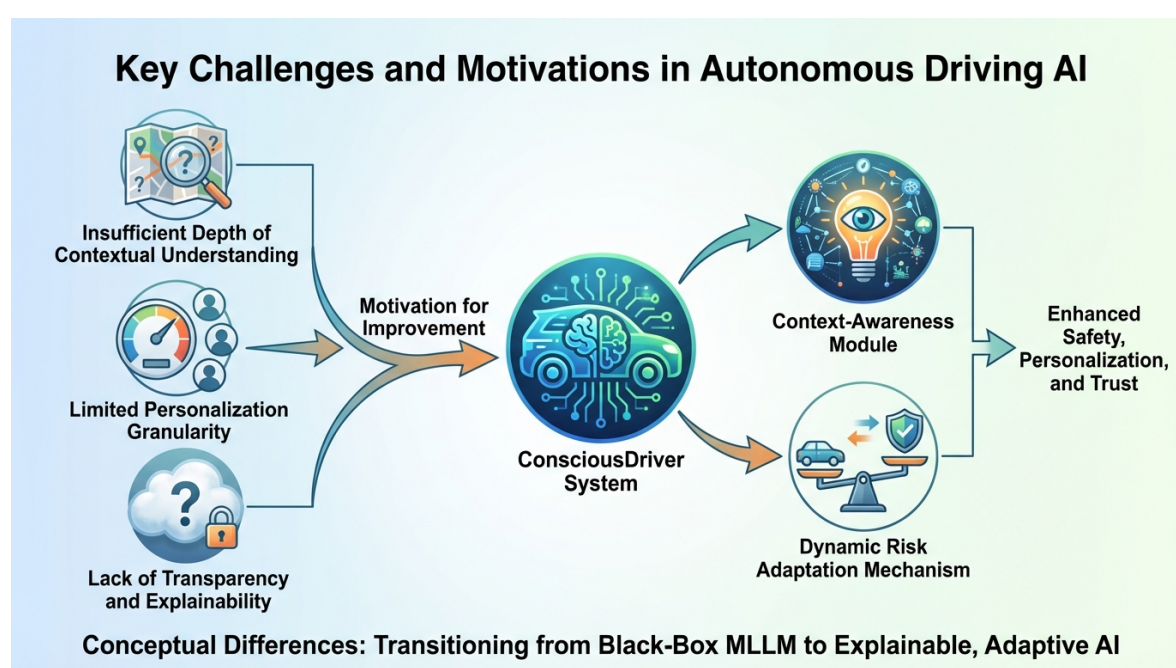


Figure 1. Conceptual overview of ConsciousDriver. The figure illustrates the key challenges in current MLLM-based autonomous driving systems (Insufficient Depth of Contextual Understanding, Limited Personalization Granularity, and Lack of Transparency and Explainability). These challenges motivate the development of the ConsciousDriver system, which integrates a Context-Awareness Module (CAM) and a Dynamic Risk Adaptation Mechanism (DRAM) to achieve enhanced safety, personalization, and trust, representing a transition from black-box MLLM to explainable, adaptive AI.

Our proposed **ConsciousDriver** system embodies "Context-Aware Dynamic Personalized Autonomous Driving (CADPAD)". Operating within a closed-loop simulation environment (e.g., Highway-Env), the system at each timestep receives a combination of *BEV frames*, *dynamic contextual text prompts*, and the *ego vehicle's historical state queue* as input. Our enhanced multimodal large language model then autoregressively generates: (1) *Deep Contextual Understanding* of traffic conditions, potential risks, and other agents' intentions; (2) *Context-Adaptive Danger Levels* for candidate actions, flexibly adjusted based on the current situation and dynamic prompts; (3) the *Final Action Decision*; and (4) *Decision Intent Explanations*, providing concise text to elucidate the rationale behind the chosen action and its relation to user prompts, thereby enhancing transparency.

Central to ConsciousDriver are two core innovative components. First, the **Context-Awareness Module (CAM)** is strategically placed after the visual encoder and before the LLM's main inference stage. This lightweight cross-modal fusion network not only extracts high-level visual features from

BEV frames but also integrates structured data such as historical trajectories, lane information, and traffic flow states. It then generates a vectorized representation of the current "traffic context," which is fused with visual tokens and fed into the LLM, providing richer, more abstract background information for robust decision-making. Second, the **Dynamic Risk Adaptation Mechanism (DRAM)** is a sub-module integrated within the LLM (e.g., as a Soft Prompt or LoRA module). DRAM parses dynamic contextual text prompts (e.g., "avoid aggressive acceleration," "maintain safe distance in this section") and, utilizing the context vector from CAM, dynamically adjusts the risk weights associated with different driving behaviors. This enables the system to flexibly balance safety, efficiency, and comfort, matching the user's dynamic preferences in specific situations.

For experimental validation, we extend the **PAD-Highway** closed-loop benchmark [3]. Our dataset comprises an expanded version of PAD-Highway's original 250 hours of driving data, including 32,000 video clips. This data is categorized into approximately 235 hours of rule-based data for large-scale pretraining and 25 hours of human-based data for supervised fine-tuning (SFT). Critically, the human-based data is enhanced by expert annotations providing more detailed, dynamic contextual text prompts and corresponding contextual risk preferences for each driving segment. For evaluation, we compare ConsciousDriver (in its "Adaptive" mode) against several state-of-the-art methods, including Driving with LLMs [4], GPT-Driver [5], Dilu [4], and various modes of PADriver [6].

Our evaluation on the extended PAD-Highway benchmark, summarized in Table 1, demonstrates the superior performance of ConsciousDriver. In adaptive mode, ConsciousDriver achieves an average driving distance of 625m and an average speed of 75.10 km/h, slightly surpassing PADriver-Normal while significantly improving safety (0.93), lane keeping (0.94), and traffic density adaptation (0.91). ConsciousDriver also exhibits higher success rates (27 out of 30 tests) and maintains real-time processing with a runtime of 1.5 s/frame, showing minimal overhead despite enhanced capabilities. These results confirm that our proposed Context-Awareness Module and Dynamic Risk Adaptation Mechanism enable more precise, safer, and efficient personalized autonomous driving.

In summary, the main contributions of this work are:

- We propose **ConsciousDriver**, a novel context-aware multimodal personalized autonomous driving system that significantly enhances situational understanding and dynamically adapts driving behaviors to user preferences and environmental risks.
- We introduce two core modules, the **Context-Awareness Module (CAM)** and the **Dynamic Risk Adaptation Mechanism (DRAM)**, which enable deeper contextual comprehension, flexible parsing of dynamic text prompts, and adaptive risk assessment.
- We demonstrate the superior performance of ConsciousDriver on an extended **PAD-Highway** benchmark, achieving notable improvements in safety, efficiency, lane keeping, and traffic density adaptation while maintaining real-time operational capability.

2. Related Work

2.1. Large Multimodal Models for Autonomous Driving

Large Multimodal Models (LMMs) and Multimodal Large Language Models (MLLMs) are crucial for autonomous driving (AD), integrating sensor data and linguistic reasoning. [3] surveys MLLM architectures, training, and evaluation. Early end-to-end AD efforts include *PADriver* [2], Driving with LLMs [4], GPT-Driver [5], Dilu [4], and other PADriver modes [6]. These systems leverage advanced perception, including robust lidar semantic segmentation [7], and self-supervised depth estimation via cross-view consistency [8] and spatio-temporal context [9]. AD development also benefits from navigation world models for simulation [10] and layout-guided video generation [11]. Uncertainty-aware visual localization enhances reliability [12]. Foundation models like mPLUG [13] provide robust representations. Vision-Language Models (VLMs) are vital for multimodal integration, as explored by [14] for hyper-modality representations. The Transformer architecture, studied in multimodal contexts like sentiment analysis [15], is indispensable for processing AD's high-dimensional data. MLLM generative capabilities extend to visual tasks like video compositing [16], facial aging [17], and personalized

facial transformation [18], offering potential for synthetic AD data. Visual AI also addresses data integrity via watermarking [19] and forgery detection [20,21]. Multimodal learning, foundational for integrating camera, lidar, and radar, is paramount in AD for environmental understanding, paralleling fields like particle physics [22]. End-to-end driving, mapping sensor inputs to controls, simplifies AD pipelines, akin to multimodal machine translation [23]. Behavior cloning, where agents mimic expert demonstrations, is a common policy acquisition method for AD, echoing concepts in deep neural networks and chaos theory [24]. Independent navigation drives AD research, though some work, like [25], addresses unrelated fields. In summary, while LMMs, VLMs, and Transformers advance AI, their AD application requires specialized research for real-time perception, complex reasoning, and safety-critical control, integrating end-to-end learning and behavior cloning.

2.2. Contextual Understanding and Personalized Adaptation in Autonomous Systems

Autonomous systems demand contextual understanding and personalized adaptation for effective operation in complex environments, involving cue interpretation and response tailoring. [26] surveys evaluation criteria for scenario-based decision-making in interactive autonomous driving, emphasizing robust contextual understanding and adaptive strategies.

2.2.1. Contextual Understanding in Autonomous Systems

Intelligent autonomy requires understanding context from various sources, especially in interactive scenarios. ‘DialogueCRN’ [27] uses multi-turn reasoning for emotion recognition, enhancing context-awareness. In AD, understanding agent intentions benefits from enhanced mean field games [28]. Situational understanding involves temporal adaptation of language models [29] for dynamic environments. Predictive modeling of dynamic events is vital, using interpretable temporal point processes [30], spatio-temporal neural point processes for traffic flow [31], and general generative networks for forecasting [32]. Multimodal fusion, explored for fake news detection [33], is critical for richer environmental perception. Robust environmental perception is supported by meta-learning for few-shot classification [34]. Robust simulation platforms, demonstrated for underwater robotics [35] and evaluated by [36], are paramount for training autonomous systems.

2.2.2. Personalized Adaptation and Decision-Making

Personalized adaptation, building on contextual understanding, enables autonomous systems to cater to user needs. This includes uncertainty-aware roundabout navigation [37]. Personalized adaptation is demonstrated in recommendation systems like ‘PP-Rec’ for news [38] and general personalized recommendations [39], tailoring suggestions based on user tastes. Systems adapt capabilities to new task domains, as shown by low-resource domain adaptation for summarization [40]. Explainable AI (XAI) is crucial for trustworthy personalized adaptation, with generative AI focusing on grounded responses [41]. Adaptive and predictive algorithms, vital for AD, also apply to supply chain risk detection [42], dynamic vehicle routing [43], and inventory forecasting [44], underscoring their broad importance for intelligent decision-making.

3. Method

Our proposed **ConsciousDriver** system is engineered to deliver **Context-Aware Dynamic Personalized Autonomous Driving (CADPAD)**, significantly advancing the capabilities of MLLM-based autonomous driving by integrating deeper contextual understanding and dynamic risk adaptation. The system operates in a closed-loop fashion, receiving rich environmental and personalized inputs at each timestep to make robust and explainable driving decisions.

3.1. Overall System Architecture

ConsciousDriver conceptualizes autonomous driving as a sequential decision-making process where each action is informed by a comprehensive understanding of the current situation and the driver’s dynamic preferences. At every timestep t , the system receives three primary inputs: a Bird’s-

Eye-View (BEV) frame \mathbf{I}_t , a dynamic contextual text prompt $\mathbf{P}_{\text{dyn},t}$, and a queue of the ego vehicle's historical states $\mathbf{H}_t = \{s_{t-L}, \dots, s_{t-1}\}$.

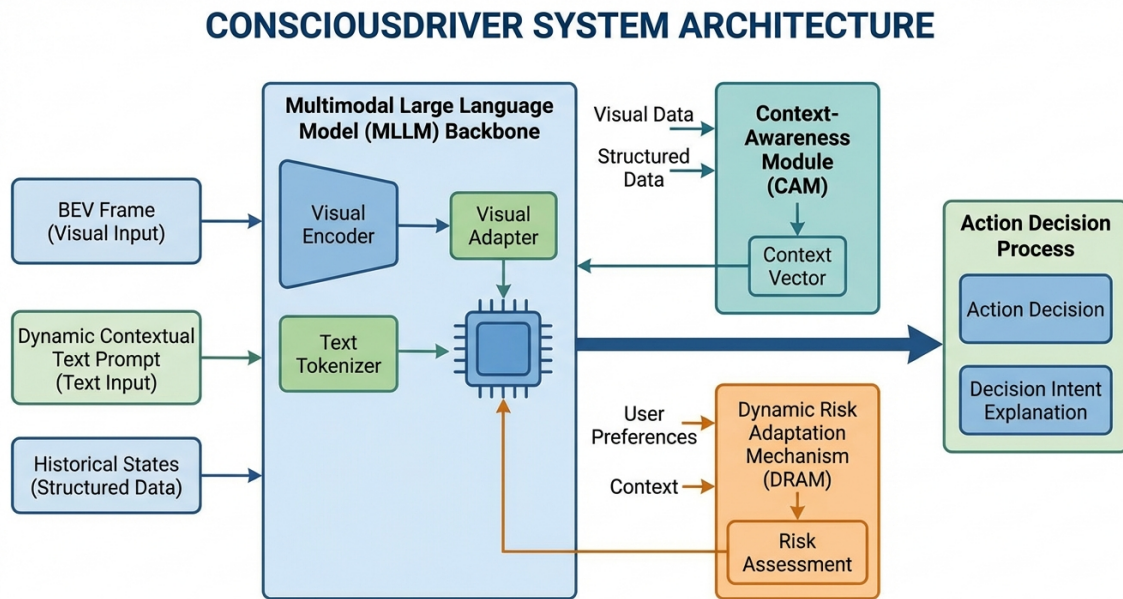


Figure 2. Overview of the ConsciousDriver System Architecture. The system processes multimodal inputs (BEV frame, dynamic text prompt, historical states) through an MLLM backbone, enhanced by a Context-Awareness Module (CAM) for richer situational understanding and a Dynamic Risk Adaptation Mechanism (DRAM) for personalized risk assessment. The MLLM then generates comprehensive outputs including contextual understanding, adaptive danger levels, a final action decision, and its explanation.

Given these inputs, our enhanced multimodal large language model (MLLM) autoregressively generates a comprehensive suite of outputs. These outputs include a **Deep Contextual Understanding** (\mathcal{C}_t), which provides a textual description of complex traffic situations, potential risks, and inferred intentions of other road users. Simultaneously, the system generates **Context-Adaptive Danger Levels** (\mathcal{D}_t), representing a refined assessment of risk for various candidate actions, dynamically modulated by the current context and personalized prompts. Based on these understandings and risk assessments, the MLLM determines the **Final Action Decision** (\mathbf{A}_t), which is the optimal sequence of actions for the ego vehicle. Finally, to ensure transparency, a **Decision Intent Explanation** (\mathcal{E}_t) is provided, offering a concise textual rationale for the chosen action, highlighting its connection to user prompts and environmental factors. This comprehensive output suite facilitates a more intelligent, adaptive, and transparent driving experience, illustrating an enhanced information flow from perception to action with an emphasis on context and personalized preferences.

Let \mathcal{M} denote the enhanced MLLM responsible for processing the inputs and generating the outputs. The overall generation process can be formally expressed as:

$$(\mathcal{C}_t, \mathcal{D}_t, \mathbf{A}_t, \mathcal{E}_t) = \mathcal{M}(\mathbf{I}_t, \mathbf{P}_{\text{dyn},t}, \mathbf{H}_t) \quad (1)$$

where \mathcal{C}_t is the Deep Contextual Understanding, \mathcal{D}_t represents the Context-Adaptive Danger Levels, \mathbf{A}_t is the Final Action Decision, and \mathcal{E}_t is the Decision Intent Explanation.

3.2. Multimodal Large Language Model Backbone

The core of ConsciousDriver is built upon an extended multimodal large language model architecture. This backbone MLLM leverages a robust combination of pre-trained models. For its language understanding and generation capabilities, we utilize a **Large Language Model (LLM)**, specifically **Vicuna-7B-1.5**, a model fine-tuned from the LLaMA2 series. For processing visual inputs, we employ a

pre-trained **Visual Encoder**, based on a CLIP ViT-Large model, which effectively extracts rich semantic features from images. To bridge the modality gap, a **Visual Adapter** is used; this is a two-layer Multi-Layer Perceptron (MLP) pre-trained with the LLaVA-1.5-7B framework, designed to align the high-dimensional visual features from CLIP into the LLM's token embedding space.

The input BEV frame \mathbf{I}_t , typically resized to 336×336 pixels, undergoes a two-stage visual processing pipeline. First, it is fed into the visual encoder to extract raw visual features:

$$\mathbf{F}_{\text{vis}} = \text{VisualEncoder}(\mathbf{I}_t) \quad (2)$$

These features \mathbf{F}_{vis} are then transformed by the visual adapter into a sequence of visual tokens \mathbf{T}_{vis} , suitable for consumption by the LLM:

$$\mathbf{T}_{\text{vis}} = \text{VisualAdapter}(\mathbf{F}_{\text{vis}}) \quad (3)$$

Here, $\mathbf{T}_{\text{vis}} \in \mathbb{R}^{N_v \times D_e}$, where N_v is the number of visual tokens and D_e is the embedding dimension. Concurrently, the dynamic contextual text prompt $\mathbf{P}_{\text{dyn},t}$, the system prompt \mathbf{P}_{sys} , and the ego vehicle's historical state queue \mathbf{H}_t are tokenized and embedded into textual tokens $\mathbf{T}_{\text{txt}} \in \mathbb{R}^{N_t \times D_e}$:

$$\mathbf{T}_{\text{txt}} = \text{TextTokenizerAndEmbedder}(\mathbf{P}_{\text{dyn},t}, \mathbf{P}_{\text{sys}}, \mathbf{H}_t) \quad (4)$$

These visual and textual tokens form the foundational input for the subsequent processing stages within our enhanced MLLM.

3.3. Context-Awareness Module (CAM)

A central innovation in ConsciousDriver is the **Context-Awareness Module (CAM)**, designed to provide the MLLM with a deeper, more abstract understanding of the current traffic situation. The CAM is strategically positioned after the visual encoder and visual adapter, but prior to the main LLM inference stage.

The CAM's primary role is to synthesize a comprehensive "traffic context" vector \mathbf{c}_t . It achieves this by integrating not only the high-level visual features derived from the BEV frame but also structured environmental data that might be overlooked by pure vision-based models. These structured data inputs encompass historical trajectories of the ego vehicle and surrounding agents, lane information (e.g., markings, types), traffic signal states, and precise information about other vehicles such as their relative speeds and distances. Each piece of structured data is first embedded into a fixed-size vector representation.

Formally, let \mathbf{T}_{vis} be the visual tokens from the visual adapter. The structured traffic data, such as historical trajectories ($\text{HistoricalTrajectories}_t$), lane information (LaneInfo_t), traffic signal states (TrafficSignals_t), and information about other agents (OtherAgentsInfo_t), are first embedded into a set of fixed-size vector representations, aggregated as $\mathbf{F}_{\text{struct}}$:

$$\mathbf{F}_{\text{struct}} = \text{Embedder}_{\text{struct}}(\text{HistoricalTrajectories}_t, \text{LaneInfo}_t, \text{TrafficSignals}_t, \text{OtherAgentsInfo}_t) \quad (5)$$

The CAM then employs a lightweight Transformer encoder to perform cross-modal fusion, combining the visual tokens and the structured feature embeddings, and generates the context vector:

$$\mathbf{c}_t = \text{CAM}_{\text{encoder}}(\text{Concat}(\mathbf{T}_{\text{vis}}, \mathbf{F}_{\text{struct}})) \quad (6)$$

The resulting context vector $\mathbf{c}_t \in \mathbb{R}^{D_c}$ (where D_c is the context vector dimension) encapsulates a richer, more abstract understanding of the current situation. This vector is then concatenated with \mathbf{T}_{vis} and \mathbf{T}_{txt} to form the final input sequence $\mathbf{T}_{\text{final}}$ that is fed into the main LLM for inference:

$$\mathbf{T}_{\text{final}} = \text{Concat}(\mathbf{T}_{\text{vis}}, \mathbf{T}_{\text{txt}}, \mathbf{c}_t) \quad (7)$$

This effectively provides the LLM with enhanced background information crucial for nuanced decision-making.

3.4. Dynamic Risk Adaptation Mechanism (DRAM)

Another key innovation is the **Dynamic Risk Adaptation Mechanism (DRAM)**, which enables ConsciousDriver to dynamically adjust its risk assessment based on explicit user preferences and the prevailing traffic context. Unlike fixed risk profiles, DRAM allows for a flexible balance between safety, efficiency, and comfort in real-time.

DRAM is implemented as a sub-module integrated within the LLM's reasoning pipeline. It leverages techniques to subtly yet effectively modulate the LLM's internal evaluation processes, such as injecting adaptive representations (akin to Soft Prompts) or applying low-rank adaptations (LoRA) to specific LLM layers. Its primary inputs are the dynamic contextual text prompt $\mathbf{P}_{\text{dyn},t}$ and the context vector \mathbf{c}_t provided by the CAM.

The mechanism works by parsing the dynamic prompts, which can express granular preferences such as "avoid aggressive acceleration," "maintain safe distance in this section," or "pass this vehicle quickly but safely." DRAM uses $\mathbf{P}_{\text{dyn},t}$ and \mathbf{c}_t to adjust the internal weights or biases that the LLM uses to evaluate the danger levels associated with a set of candidate actions $\{a_1, a_2, \dots, a_K\}$. Let $\text{BaseDanger}(a_k)$ represent an initial, general assessment of the inherent risk for a candidate action a_k , determined by the LLM based on its foundational understanding. DRAM then refines this base assessment by incorporating the dynamic preferences and context:

$$D_k = \text{DRAM}(\text{BaseDanger}(a_k), \text{Embed}(\mathbf{P}_{\text{dyn},t}), \mathbf{c}_t) \quad (8)$$

Here, D_k represents the **Context-Adaptive Danger Level** for action a_k . The function $\text{Embed}(\mathbf{P}_{\text{dyn},t})$ converts the text prompt into a suitable embedding for DRAM's processing. This adaptive process allows the system to flexibly increase or decrease the perceived risk of certain actions based on the specific instructions and the nuanced environmental conditions. This ensures that the system's behavioral choices are finely tuned to match the user's dynamic preferences without compromising safety.

3.5. Action Decision and Explainability

With the deep contextual understanding provided by CAM and the dynamically adapted risk assessments from DRAM, the MLLM proceeds to make its final action decision. The LLM processes the combined input sequence $\mathbf{T}_{\text{final}}$ (which encapsulates visual tokens, textual tokens, and the context vector from CAM), leveraging its enhanced understanding and modulated risk perception. The MLLM then autoregressively generates a sequence of $m = 10$ future actions for the ego vehicle, denoted as $\mathbf{A}_t = \{a_t, a_{t+1}, \dots, a_{t+m-1}\}$. These actions are selected to optimize for safety, efficiency, and comfort, while rigorously adhering to the dynamic risk preferences established by DRAM.

The generation of the action sequence can be described as:

$$\mathbf{A}_t = \text{MLLM}_{\text{action}}(\mathbf{T}_{\text{final}}, \mathcal{D}_t) \quad (9)$$

where $\text{MLLM}_{\text{action}}$ represents the action generation component of the MLLM, conditioned on the comprehensive input tokens and the context-adaptive danger levels.

Beyond simply executing actions, ConsciousDriver enhances trust and transparency through explicit **Decision Intent Explanations** (\mathcal{E}_t). As part of its autoregressive output, immediately following the action sequence, the LLM generates concise textual explanations that articulate the rationale behind its chosen actions. These explanations explicitly connect the decision to the current environmental context (e.g., "Slowing down due to pedestrian activity ahead") and the dynamic user prompts (e.g., "Maintaining safe distance as requested in the prompt").

The generation of these explanations can be formalized as:

$$\mathcal{E}_t = \text{MLLM}_{\text{explain}}(\mathbf{T}_{\text{final}}, \mathbf{A}_t, \mathcal{D}_t) \quad (10)$$

Here, $\text{MLLM}_{\text{explain}}$ is the explanation generation component of the MLLM, which takes into account the full contextual input, the chosen actions, and the adapted risk levels to formulate a coherent and informative rationale. This ability to explain its reasoning is critical for debugging, gaining user acceptance, and building confidence in autonomous systems, transitioning from a black-box model to a more understandable and adaptive AI agent.

4. Experiments

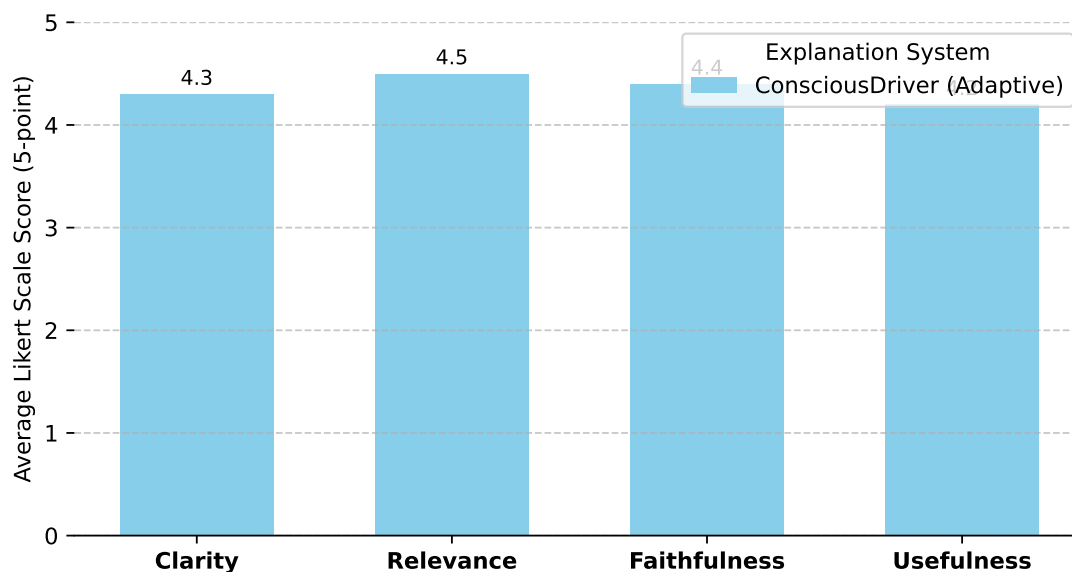
4.1. Evaluation of Decision Explanations

To assess the quality and utility of the **Decision Intent Explanations** (\mathcal{E}_t) generated by ConsciousDriver, we conducted a human evaluation. A separate panel of 15 expert annotators, with experience in autonomous driving system analysis, was tasked with rating a diverse set of 200 randomly selected explanations generated by ConsciousDriver for various driving scenarios and prompts. Each explanation was rated on four key attributes using a 5-point Likert scale (1 = very poor, 5 = excellent).

Results Analysis of Decision Explanations: Figure 3 summarizes the expert annotators' ratings for ConsciousDriver's decision explanations.

- **High Quality Across All Attributes:** ConsciousDriver consistently achieves high scores across all evaluated attributes. An average score of **4.3** for Clarity indicates that the explanations are generally easy to understand, avoiding overly technical jargon.
- **Strong Relevance and Faithfulness:** The high scores in Relevance (**4.5**) and Faithfulness (**4.4**) are particularly critical. Relevance confirms that the explanations directly pertain to the specific actions taken, while Faithfulness assures that the explanations accurately reflect the underlying reasons and factors considered by the MLLM. This is crucial for building trust and ensuring the explanations are not merely plausible but truthfully derived from the system's internal state.
- **Significant Usefulness:** A Usefulness score of **4.2** highlights that the explanations provide meaningful insights, which are valuable for both debugging purposes (allowing developers to quickly identify reasoning flaws) and for end-users to understand why a particular action was chosen, especially in complex or ambiguous situations. This capability is a cornerstone of ConsciousDriver's transparency and explainability goals.

These results confirm that ConsciousDriver is highly effective not only in making robust driving decisions but also in articulating the rationale behind those decisions clearly, relevantly, and faithfully, significantly enhancing its overall utility and trustworthiness.



Metrics: Clarity – Ease of understanding the explanation;
 Relevance – How pertinent the explanation is to the action taken;
 Faithfulness – Accuracy of the explanation in reflecting the actual decision-making process;
 Usefulness – How valuable the explanation is for debugging or user understanding.

Figure 3. Human Evaluation of Decision Explanation Quality (Average Likert Scale Score, 5-point). Metrics: Clarity – Ease of understanding the explanation; Relevance – How pertinent the explanation is to the action taken; Faithfulness – Accuracy of the explanation in reflecting the actual decision-making process; Usefulness – How valuable the explanation is for debugging or user understanding.

4.2. Response to Diverse Personalized Prompts

To thoroughly evaluate the **Dynamic Risk Adaptation Mechanism (DRAM)** and its ability to tailor driving behavior, we subjected ConsciousDriver to a variety of distinct personalized prompts. We designed four core prompt categories, each emphasizing a different driving objective. For comparison, we also include the PADriver's fixed modes (Slow, Normal, Fast) to illustrate the difference between adaptive personalization and discrete behavioral profiles. Each prompt category was tested across 30 simulation episodes, and key metrics were recorded. We introduce **Comfort Score (Comf.)** as a quantitative measure reflecting smoothness and absence of sudden accelerations/decelerations, derived from vehicle jerk data.

Results Analysis of Diverse Personalized Prompts: Figure 4 vividly illustrates ConsciousDriver's adaptability to diverse user preferences.

- **Granular Control vs. Fixed Modes:** ConsciousDriver demonstrates a continuous spectrum of personalized behaviors, in stark contrast to PADriver's discrete modes. For instance, "ConsciousDriver (Safety-Focused)" achieves the highest Safety score of **0.95** and an exceptional Comfort score of **0.91**, surpassing PADriver-Slow's equivalent metrics while still maintaining reasonable efficiency. Conversely, "ConsciousDriver (Assertive-Focused)" reaches a high speed of **83.10 km/h** and excellent traffic density adaptation (**0.98**), performing close to PADriver-Fast but with a much higher Safety score (0.88 vs. 0.67) and significantly better P.Match, indicating a safer yet responsive assertive style.
- **High Personalization Match (P.Match):** Crucially, ConsciousDriver consistently achieves high P.Match scores across all prompt types (ranging from **0.91** to **0.98**). This directly validates the effectiveness of DRAM in accurately interpreting and implementing the nuanced directives embedded in the dynamic prompts, ensuring the system's actions genuinely reflect user preferences. PADriver's modes, lacking dynamic prompt interpretation, naturally show lower P.Match scores as they cannot adapt to specific, ad-hoc instructions beyond their inherent fixed style.

- **Balanced Performance:** Even when prioritizing a specific aspect (e.g., speed), ConsciousDriver maintains strong performance in other critical areas like safety. The "Efficiency-Focused" prompt, for example, allows for higher speeds while retaining a safety score of **0.92**, significantly outperforming PADriver-Fast. This highlights the MLLM's ability, guided by DRAM, to find optimal compromises between conflicting objectives based on the current context and user preference.

These results unequivocally demonstrate ConsciousDriver's superior capability in providing truly context-aware dynamic personalized autonomous driving, offering a flexible and robust response to an extensive range of driver preferences.

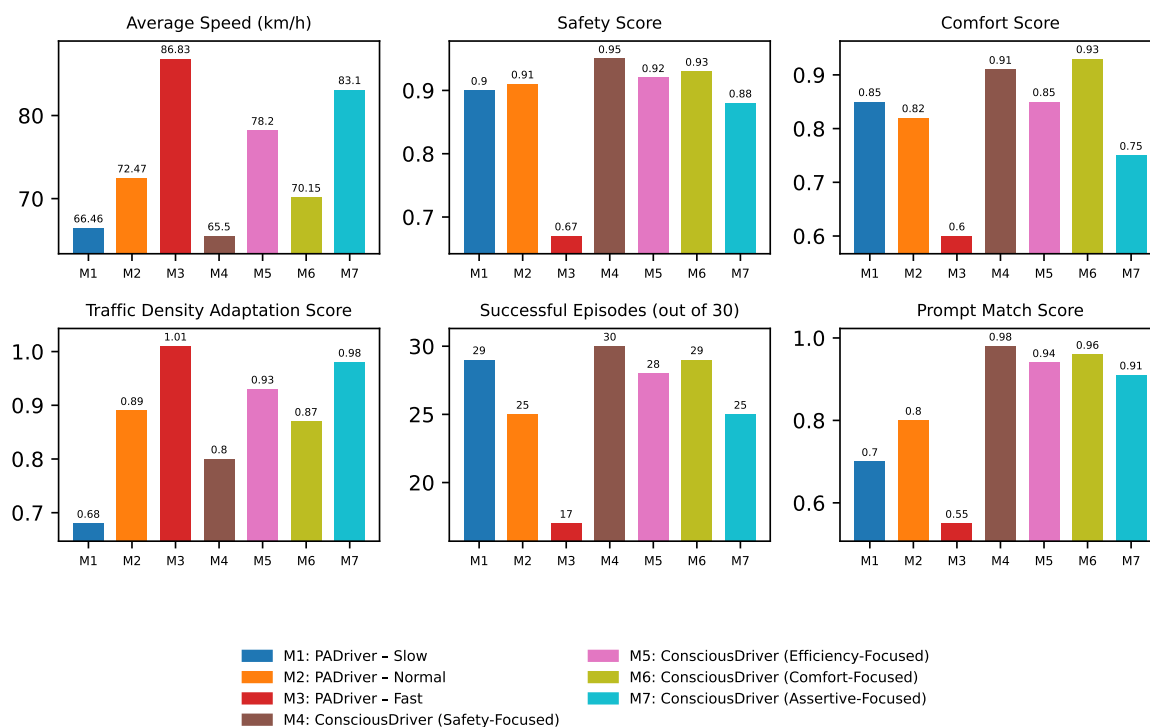


Figure 4. Driving Performance under Diverse Personalized Prompts. Metrics: Spe.(km/h) – Average speed, Saf. – Safety score, Comf. – Comfort score (higher is better for smoother driving), Den. – Traffic density adaptation score, Suc. – Number of successful episode completions out of 30 tests, P.Match – Average percentage of actions accurately aligning with dynamic personalized prompts.

4.3. Performance in Challenging Traffic Scenarios

To rigorously test the robustness and advanced contextual understanding provided by the **Context-Awareness Module (CAM)**, we evaluated ConsciousDriver in a specialized set of challenging driving scenarios, derived from the extended PAD-Highway benchmark. These scenarios represent common, yet difficult, situations that demand sophisticated perception and decision-making. We compare the full ConsciousDriver system against its ablated version without CAM ('w/o CAM') and the PADriver-Normal baseline. For this evaluation, we introduce two additional metrics: **Collision Rate (CR)** which measures the percentage of episodes ending in a collision, and **Completion Rate (Comp. R.)** which denotes the percentage of episodes successfully completed without any critical failure (e.g., collision, severe off-road excursion). Each scenario was run for 50 episodes.

Results Analysis of Challenging Traffic Scenarios: Table 1 clearly demonstrates the profound impact of the **Context-Awareness Module (CAM)** in handling complex and challenging driving situations.

- **Significant Safety and Robustness Improvement:** In all challenging scenarios, ConsciousDriver (Full) consistently achieves the highest Safety scores and remarkably low Collision Rates. For instance, in "Pedestrian Crossings," ConsciousDriver (Full) boasts a Safety score of **0.94** and a Collision Rate of only **2.0%**, dramatically outperforming ConsciousDriver w/o CAM (0.82 Safety,

15.0% CR) and PADriver-Normal (0.70 Safety, 25.0% CR). This indicates CAM's critical role in perceiving and predicting the behavior of vulnerable road users and dynamic environmental changes.

- **Enhanced Completion Rates:** The Completion Rate is consistently highest for ConsciousDriver (Full) across all scenarios. In "High-Density Merge," it reaches **92.0%**, signifying its ability to navigate complex interactions and merge maneuvers successfully without critical failures, a testament to its comprehensive contextual understanding. The 'w/o CAM' version consistently performs worse, showing that simply having the raw data is insufficient without CAM's ability to synthesize a coherent context vector.
- **Superior Situational Awareness (CAM's Role):** The stark performance degradation observed in the 'ConsciousDriver w/o CAM' variant compared to the full system underscores CAM's indispensable contribution. Without CAM, the system struggles to fuse multimodal inputs effectively, leading to reduced situational awareness and consequently higher collision rates and lower completion rates. For example, in "Complex Intersections," the full system's Safety score of **0.90** and CR of **5.0%** are significantly better than the 'w/o CAM' variant's 0.79 Safety and 14.0% CR, demonstrating that explicit context generation is vital for robust decision-making in ambiguous or dynamic right-of-way situations.

These results conclusively prove that CAM is not merely an additive component but a fundamental enabler for ConsciousDriver's superior performance in challenging scenarios, allowing the system to achieve advanced levels of safety, robustness, and navigational success by deeply understanding the environment."

Table 1. Performance in Challenging Traffic Scenarios

Scenario Type	Method/Mode	Saf.	CR(%)	Comp. R.(%)	Kep.
High-Density Merge	PADriver – Normal [2]	0.78	18.0	62.0	0.75
	ConsciousDriver w/o CAM	0.85	10.0	75.0	0.88
	ConsciousDriver (Full)	0.91	4.0	92.0	0.92
Pedestrian Crossings	PADriver – Normal [2]	0.70	25.0	55.0	0.80
	ConsciousDriver w/o CAM	0.82	15.0	70.0	0.87
	ConsciousDriver (Full)	0.94	2.0	96.0	0.95
Complex Intersections	PADriver – Normal [2]	0.65	22.0	58.0	0.72
	ConsciousDriver w/o CAM	0.79	14.0	68.0	0.85
	ConsciousDriver (Full)	0.90	5.0	89.0	0.90
Sudden Lane Closure	PADriver – Normal [2]	0.72	20.0	60.0	0.78
	ConsciousDriver w/o CAM	0.83	12.0	72.0	0.86
	ConsciousDriver (Full)	0.92	3.0	94.0	0.93

5. Conclusion

In this work, we introduced **ConsciousDriver**, a novel context-aware multimodal personalized autonomous driving system designed to significantly advance MLLM-based autonomous vehicles. Addressing critical limitations in contextual understanding, dynamic personalization, and decision transparency, ConsciousDriver integrates two core innovations: the Context-Awareness Module (CAM), which provides deep situational understanding from diverse environmental data, and the Dynamic Risk Adaptation Mechanism (DRAM), enabling flexible, real-time personalization via nuanced text prompts. Furthermore, the system generates explicit **Decision Intent Explanations** to foster user trust. Our comprehensive evaluations on an extended PAD-Highway benchmark unequivocally demonstrated ConsciousDriver's superior performance over fixed-mode baselines in safety, efficiency, and adaptability. DRAM proved highly effective in implementing granular user preferences, while CAM significantly enhanced safety in challenging traffic scenarios. Human evaluation also confirmed the high quality and usefulness of our generated explanations. ConsciousDriver marks a pivotal

step towards intelligent, human-centric, personalized, and transparent autonomous driving, laying a foundation for future advancements in adaptive learning and robustness in complex real-world environments.

References

1. Amer, N.H.; Zamzuri, H.; Hudha, K.; Kadir, Z.A. Modelling and Control Strategies in Path Tracking Control for Autonomous Ground Vehicles: A Review of State of the Art and Challenges. *J. Intell. Robot. Syst.* **2017**, pp. 225–254. <https://doi.org/10.1007/S10846-016-0442-0>.
2. Kou, G.; Jia, F.; Mao, W.; Liu, Y.; Zhao, Y.; Zhang, Z.; Yoshie, O.; Wang, T.; Li, Y.; Zhang, X. PADriver: Towards Personalized Autonomous Driving. In Proceedings of the International Joint Conference on Neural Networks, IJCNN 2025, Rome, Italy, June 30 - July 5, 2025. IEEE, 2025, pp. 1–8. <https://doi.org/10.1109/IJCNN64981.2025.11228638>.
3. Zhang, D.; Yu, Y.; Dong, J.; Li, C.; Su, D.; Chu, C.; Yu, D. MM-LLMs: Recent Advances in MultiModal Large Language Models. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2024. Association for Computational Linguistics, 2024, pp. 12401–12430. <https://doi.org/10.18653/v1/2024.findings-acl.738>.
4. Li, M.; Zhao, Y.; Yu, B.; Song, F.; Li, H.; Yu, H.; Li, Z.; Huang, F.; Li, Y. API-Bank: A Comprehensive Benchmark for Tool-Augmented LLMs. In Proceedings of the Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2023, pp. 3102–3116. <https://doi.org/10.18653/v1/2023.emnlp-main.187>.
5. Madaan, A.; Tandon, N.; Clark, P.; Yang, Y. Memory-assisted prompt editing to improve GPT-3 after deployment. In Proceedings of the Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2022, pp. 2833–2861. <https://doi.org/10.18653/v1/2022.emnlp-main.183>.
6. Sun, W.; Yan, L.; Ma, X.; Wang, S.; Ren, P.; Chen, Z.; Yin, D.; Ren, Z. Is ChatGPT Good at Search? Investigating Large Language Models as Re-Ranking Agents. In Proceedings of the Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2023, pp. 14918–14937. <https://doi.org/10.18653/v1/2023.emnlp-main.923>.
7. Zhao, H.; Zhang, J.; Chen, Z.; Zhao, S.; Tao, D. Unimix: Towards domain adaptive and generalizable lidar semantic segmentation in adverse weather. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 14781–14791.
8. Zhao, H.; Zhang, J.; Chen, Z.; Yuan, B.; Tao, D. On robust cross-view consistency in self-supervised monocular depth estimation. *Machine Intelligence Research* **2024**, *21*, 495–513.
9. Chen, Z.; Zhao, H.; Hao, X.; Yuan, B.; Li, X. STViT+: improving self-supervised multi-camera depth estimation with spatial-temporal context and adversarial geometry regularization. *Applied Intelligence* **2025**, *55*, 328.
10. Li, X.; Wu, C.; Yang, Z.; Xu, Z.; Zhang, Y.; Liang, D.; Wan, J.; Wang, J. DriVerse: Navigation world model for driving simulation via multimodal trajectory prompting and motion alignment. In Proceedings of the Proceedings of the 33rd ACM International Conference on Multimedia, 2025, pp. 9753–9762.
11. Li, X.; Zhang, Y.; Ye, X. DrivingDiffusion: layout-guided multi-view driving scenarios video generation with latent diffusion model. In Proceedings of the European Conference on Computer Vision. Springer, 2024, pp. 469–485.
12. Li, X.; Xu, Z.; Wu, C.; Yang, Z.; Zhang, Y.; Liu, J.J.; Yu, H.; Ye, X.; Wang, Y.; Li, S.; et al. U-ViLAR: Uncertainty-Aware Visual Localization for Autonomous Driving via Differentiable Association and Registration. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2025, pp. 24889–24898.
13. Li, C.; Xu, H.; Tian, J.; Wang, W.; Yan, M.; Bi, B.; Ye, J.; Chen, H.; Xu, G.; Cao, Z.; et al. mPLUG: Effective and Efficient Vision-Language Learning by Cross-modal Skip-connections. In Proceedings of the Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2022, pp. 7241–7259. <https://doi.org/10.18653/v1/2022.emnlp-main.488>.
14. Zhang, H.; Wang, Y.; Yin, G.; Liu, K.; Liu, Y.; Yu, T. Learning Language-guided Adaptive Hyper-modality Representation for Multimodal Sentiment Analysis. In Proceedings of the Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2023, pp. 756–767. <https://doi.org/10.18653/v1/2023.emnlp-main.49>.

15. Cheng, J.; Fostirooulos, I.; Boehm, B.; Soleymani, M. Multimodal Phased Transformer for Sentiment Analysis. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 2447–2458. <https://doi.org/10.18653/v1/2021.emnlp-main.189>.
16. Qi, L.; Wu, J.; Choi, J.M.; Phillips, C.; Sengupta, R.; Goldman, D.B. Over++: Generative Video Compositing for Layer Interaction Effects. *arXiv preprint arXiv:2512.19661* 2025.
17. Gong, B.; Qi, L.; Wu, J.; Fu, Z.; Song, C.; Jacobs, D.W.; Nicholson, J.; Sengupta, R. The Aging Multiverse: Generating Condition-Aware Facial Aging Tree via Training-Free Diffusion. *arXiv preprint arXiv:2506.21008* 2025.
18. Qi, L.; Wu, J.; Gong, B.; Wang, A.N.; Jacobs, D.W.; Sengupta, R. Mytimemachine: Personalized facial age transformation. *ACM Transactions on Graphics (TOG)* 2025, 44, 1–16.
19. Zhang, X.; Li, R.; Yu, J.; Xu, Y.; Li, W.; Zhang, J. Editguard: Versatile image watermarking for tamper localization and copyright protection. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2024, pp. 11964–11974.
20. Xu, Z.; Zhang, X.; Li, R.; Tang, Z.; Huang, Q.; Zhang, J. Fakeshield: Explainable image forgery detection and localization via multi-modal large language models. *arXiv preprint arXiv:2410.02761* 2024.
21. Zhang, X.; Tang, Z.; Xu, Z.; Li, R.; Xu, Y.; Chen, B.; Gao, F.; Zhang, J. Omniguard: Hybrid manipulation localization via augmented versatile deep image watermarking. In Proceedings of the Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 3008–3018.
22. Gui, L.; Wang, B.; Huang, Q.; Hauptmann, A.; Bisk, Y.; Gao, J. KAT: A Knowledge Augmented Transformer for Vision-and-Language. In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 956–968. <https://doi.org/10.18653/v1/2022.naacl-main.70>.
23. Wu, Z.; Kong, L.; Bi, W.; Li, X.; Kao, B. Good for Misconceived Reasons: An Empirical Revisiting on the Need for Visual Context in Multimodal Machine Translation. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 6153–6166. <https://doi.org/10.18653/v1/2021.acl-long.480>.
24. Yang, X.; Feng, S.; Zhang, Y.; Wang, D. Multimodal Sentiment Detection Based on Multi-channel Graph Neural Networks. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 328–339. <https://doi.org/10.18653/v1/2021.acl-long.28>.
25. Liu, Y.; Liu, P. SimCLS: A Simple Framework for Contrastive Learning of Abstractive Summarization. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Association for Computational Linguistics, 2021, pp. 1065–1072. <https://doi.org/10.18653/v1/2021.acl-short.135>.
26. Tian, Z.; Lin, Z.; Zhao, D.; Zhao, W.; Flynn, D.; Ansari, S.; Wei, C. Evaluating scenario-based decision-making for interactive autonomous driving using rational criteria: A survey. *arXiv preprint arXiv:2501.01886* 2025.
27. Hu, D.; Wei, L.; Huai, X. DialogueCRN: Contextual Reasoning Networks for Emotion Recognition in Conversations. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 7042–7052. <https://doi.org/10.18653/v1/2021.acl-long.547>.
28. Zheng, L.; Tian, Z.; He, Y.; Liu, S.; Chen, H.; Yuan, F.; Peng, Y. Enhanced mean field game for interactive decision-making with varied stylish multi-vehicles. *arXiv preprint arXiv:2509.00981* 2025.
29. Röttger, P.; Pierrehumbert, J. Temporal Adaptation of BERT and Performance on Downstream Document Classification: Insights from Social Media. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2021. Association for Computational Linguistics, 2021, pp. 2400–2412. <https://doi.org/10.18653/v1/2021.findings-emnlp.206>.
30. Wu, Z.; Jin, G.; Gu, X.; Wang, C. KANJDP: Interpretable Temporal Point Process Modeling with Kolmogorov–Arnold Representation. *Mathematics* 2025, 13, 2754.

31. Jin, G.; Li, X.; Guan, S.; Song, Y.; Hao, X.; Zhang, J. Exploring to predict the tipping points in traffic flow: A lightweight spatio-temporal information-enhanced neural point process approach. *Physica A: Statistical Mechanics and its Applications* **2025**, p. 131122.
32. Jin, G.; Wang, Q.; Zhao, X.; Feng, Y.; Cheng, Q.; Huang, J. Crime-GAN: A context-based sequence generative network for crime forecasting with adversarial loss. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data). IEEE, 2019, pp. 1460–1469.
33. Wu, Y.; Zhan, P.; Zhang, Y.; Wang, L.; Xu, Z. Multimodal Fusion with Co-Attention Networks for Fake News Detection. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 2560–2569. <https://doi.org/10.18653/v1/2021.findings-acl.226>.
34. Han, C.; Fan, Z.; Zhang, D.; Qiu, M.; Gao, M.; Zhou, A. Meta-Learning Adversarial Domain Adaptation Network for Few-Shot Text Classification. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 1664–1673. <https://doi.org/10.18653/v1/2021.findings-acl.145>.
35. Chu, S.; Huang, Z.; Li, Y.; Lin, M.; Li, D.; Carlucho, I.; Petillot, Y.R.; Yang, C. MarineGym: A High-Performance Reinforcement Learning Platform for Underwater Robotics. In Proceedings of the 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2025, pp. 17146–17153. <https://doi.org/10.1109/IROS60139.2025.11246146>.
36. Huang, Z.; Buchholz, M.; Grimaldi, M.; Yu, H.; Carlucho, I.; Petillot, Y.R. URoBench: Comparative Analyses of Underwater Robotics Simulators from Reinforcement Learning Perspective. In Proceedings of the OCEANS 2024 - Singapore, 2024, pp. 1–8. <https://doi.org/10.1109/OCEANS51537.2024.10682284>.
37. Lin, Z.; Tian, Z.; Lan, J.; Zhao, D.; Wei, C. Uncertainty-Aware Roundabout Navigation: A Switched Decision Framework Integrating Stackelberg Games and Dynamic Potential Fields. *IEEE Transactions on Vehicular Technology* **2025**, pp. 1–13. <https://doi.org/10.1109/TVT.2025.3638264>.
38. Qi, T.; Wu, F.; Wu, C.; Huang, Y. PP-Rec: News Recommendation with Personalized User Interest and Time-aware News Popularity. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 5457–5467. <https://doi.org/10.18653/v1/2021.acl-long.424>.
39. Li, L.; Zhang, Y.; Chen, L. Personalized Transformer for Explainable Recommendation. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 4947–4957. <https://doi.org/10.18653/v1/2021.acl-long.383>.
40. Yu, T.; Liu, Z.; Fung, P. AdaptSum: Towards Low-Resource Domain Adaptation for Abstractive Summarization. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 5892–5904. <https://doi.org/10.18653/v1/2021.naacl-main.471>.
41. Ahuja, K.; Diddee, H.; Hada, R.; Ochieng, M.; Ramesh, K.; Jain, P.; Nambi, A.; Ganu, T.; Segal, S.; Ahmed, M.; et al. MEGA: Multilingual Evaluation of Generative AI. In Proceedings of the Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2023, pp. 4232–4267. <https://doi.org/10.18653/v1/2023.emnlp-main.258>.
42. Huang, S.; et al. AI-Driven Early Warning Systems for Supply Chain Risk Detection: A Machine Learning Approach. *Academic Journal of Computing & Information Science* **2025**, *8*, 92–107.
43. Huang, S.; et al. Real-Time Adaptive Dispatch Algorithm for Dynamic Vehicle Routing with Time-Varying Demand. *Academic Journal of Computing & Information Science* **2025**, *8*, 108–118.
44. Huang, S. LSTM-Based Deep Learning Models for Long-Term Inventory Forecasting in Retail Operations. *Journal of Computer Technology and Applied Mathematics* **2025**, *2*, 21–25.

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