

---

# ACCIDENTGPT: ACCIDENT ANALYSIS AND PREVENTION FROM V2X ENVIRONMENTAL PERCEPTION WITH MULTI-MODAL LARGE MODEL

---

*Lening Wang*<sup>a,b</sup>, *Han Jiang*<sup>a</sup>, *Pinlong Cai*<sup>b</sup>, *Daocheng Fu*<sup>b</sup>, *Tianqi Wang*<sup>c</sup>, *Zhiyong Cui*<sup>\* a</sup>,  
*Yilong Ren*<sup>\* a</sup>, *Haiyang Yu*<sup>\* a</sup>, *Xuesong Wang*<sup>d</sup>, and *Yinhai Wang*<sup>f</sup>

<sup>a</sup>State Key Lab of Intelligent Transportation SystemBeihang University, Beijing, China, 100191  
{leningwang; buaajh; zhiyongc; yilongren; hyyu} @buaa.edu.cn

<sup>b</sup>Shanghai Artificial Intelligence Laboratory, Shanghai, China, 200232  
{caipinlong; fudaocheng} @pjlab.org.cn

<sup>c</sup>The University of Hong Kong, Hong Kong, China, 999077  
wangtq@connect.hku.hk

<sup>d</sup>Tongji University, Shanghai, China, 201804  
wangxs@tongji.edu.cn

<sup>f</sup>University of Washington, Seattle, United States, 98195-2700  
yinhai@uw.edu

\*Corresponding author

## ABSTRACT

Traffic accidents, being a significant contributor to both human casualties and property damage, have long been a focal point of research for many scholars in the field of traffic safety. However, previous studies, whether focusing on static environmental assessments or dynamic driving analyses, as well as pre-accident predictions or post-accident rule analyses, have typically been conducted in isolation. There has been a lack of an effective framework for developing a comprehensive understanding and application of traffic safety. To address this gap, this paper introduces AccidentGPT, a comprehensive accident analysis and prevention multi-modal large model. AccidentGPT establishes a multi-modal information interaction framework grounded in multi-sensor perception, thereby enabling a holistic approach to accident analysis and prevention in the field of traffic safety. Specifically, our capabilities can be categorized as follows: for autonomous driving vehicles, we provide comprehensive environmental perception and understanding to control the vehicle and avoid collisions. For human-driven vehicles, we offer proactive long-range safety warnings and blind-spot alerts while also providing safety driving recommendations and behavioral norms through human-machine dialogue and interaction. Additionally, for traffic police and management agencies, our framework supports intelligent and real-time analysis of traffic safety, encompassing pedestrian, vehicles, roads, and the environment through collaborative perception from multiple vehicles and road testing devices. The system is also capable of providing a thorough analysis of accident causes and liability after vehicle collisions. Our framework stands as the first large model to integrate comprehensive scene understanding into traffic safety studies. Project page: <https://accidentgpt.github.io>

**Keywords** traffic accident analysis and safety · multi-modal large model · collaborative perception · human-machine interaction

## 1 Introduction

Like the majority of research in traffic safety, the study of accident analysis and prevention, aiming to enhance the safety of the traffic system and reduce casualties and property losses, is a necessary endeavor [1, 2, 3]. In previous

research, scholars have approached traffic safety from various perspectives, categorizing their studies according to traffic scenarios, events, and degrees of traffic intelligence. These investigations broadly encompass a range of focal areas: analysis of static environments [4, 5], examination of dynamic object states [6], forecasting and analysis of accidents and their causes, evaluation of human driving safety [7, 8], and in-depth analysis of comprehensive safety in autonomous driving [9, 10]. This multifaceted approach reflects the complexity and breadth of traffic safety research. For instance, in the realm of static environment analysis, a thorough examination is conducted to understand the impact of traffic safety within the broader context of road and environmental scenarios. In dynamic object state analysis, the emphasis lies on aspects like collision and conflict prevention, the study of driving behaviors, and the modeling of vehicle dynamics. Several researchers focus on accidents, engaging in proactive collision prediction before incidents and in-depth causation analysis afterwards, with the aim to reduce the frequency and severity of traffic accidents. Concurrently, as autonomous driving research rapidly advances, the focus has shifted from traditional manual driving safety towards investigating the safety of human-machine hybrid driving [11] transitions and the coexistence of mixed vehicle types. This shift aims to ensure precision in safety and control mechanisms for fully autonomous vehicles.

However, despite a common overarching goal of fostering a safer, more comprehensive traffic safety environment and preventing accidents, past research efforts have often been compartmentalized. This resulted in a fragmentation of efforts, with a notable absence of effective coordination between different tasks and modules. Consequently, there has been a lack of a unified, integrated framework capable of achieving a holistic understanding and practical application of traffic safety principles.

With the swift progression of artificial intelligence, the integration of collaborative perception [12, 13, 14] and Large Language Model (LLM) [15, 16, 17] promises a deeper understanding and enriched interaction in traffic safety. To this end, we introduce AccidentGPT, an advanced multi-modal large model designed for accident analysis and prevention, rooted in Environmental Perception. This framework is meticulously engineered to integrate and interpret a wide array of sensor data, forming a multimodal platform for interaction and comprehension. Its primary objective is to address accident analysis and prevention, smoothing the shift from human-operated driving to intelligent vehicle safety in autonomous driving.

Specifically, AccidentGPT establishes the V2X-perception framework for exhaustive scene perception and comprehension, complemented by the GPT-reasoning framework for multi-modal interaction across various specialized tasks. This leads to a holistic approach in traffic safety research. The Vehicle-to-Everything (V2X) perception framework amalgamates panoramic imagery from multiple vehicles and road-testing devices, enabling 3D object detection, Bird’s Eye View (BEV) perspective perception, and the recognition and construction of comprehensive scenarios involving pedestrian, vehicles, roads, and the environment. This approach not only augments perception and planning for autonomous driving as well as achieves an all-encompassing understanding of traffic safety scenes. It aids in real-time assessment of the traffic safety environment, alongside accident prediction and analysis. Moreover, within the GPT-reasoning framework, we draw from the innate cognition of human driving and extensive traffic safety analysis. This structure deeply integrates a LLM, merging cognitive reasoning with perceptual skills. It incorporates various traffic safety tasks into its corpus, enabling profound interactions in natural language reasoning and perception tasks.

Our contributions can be mainly divided into the following parts:

- **Autonomous Vehicle Environmental Perception:** We have developed a comprehensive scene perception and prediction system for fully autonomous vehicles. This system uses cameras on multiple vehicles and road-testing devices to create a unified BEV space, enabling 3D object detection, perspective analysis, and motion prediction. Our approach integrates multi-vehicle collaborative perception for enhanced environmental understanding and collision avoidance.
- **Safety Enhancements for Human-driven Vehicles:** For human-driven vehicles, we offer advanced safety features such as proactive remote safety warnings and blind spot alerts. Our system comprises distinct agents, labeling, and semantic modules that process visual data, including scene, BEV images, and 3D objects detection results. This multi-level modular input structure, supported by a corporate database and priority sampling module, improves the versatility of safety measures for human-driven vehicles.
- **Traffic Safety Analysis for Enforcement Agencies:** Our framework serves traffic police and management agencies by providing real-time intelligent analysis of traffic safety factors – pedestrian, vehicles, roads, and the environment. Utilizing collaborative perception from multiple vehicles and road testing devices, it delivers comprehensive reports on accident causes and responsibilities. The system’s effectiveness is enhanced by a rich pre-training on an efficient traffic safety corpus, ensuring accurate analysis and insights.

These points encapsulate the primary contributions of our capabilities, highlighting our advancements in autonomous vehicle technology, safety features for human-driven vehicles, and traffic safety analysis tools for enforcement agencies.

In summary, our framework supports all-around traffic safety research and is the first to integrate comprehensive scene understanding into traffic safety research using a large model.

## 2 Related Work

Our research is situated at the intersection of traffic safety analysis and warning tasks, autonomous driving perception and planning tasks, and the application of LLMs in the field of transportation and vehicles. Therefore, we will separately introduce the mainstream research in these three tasks and the points of integration in our research.

### 2.1 Traffic safety analysis and accident warning

According to the World Health Organization, approximately 1.3 million people die annually due to road traffic collisions [1]. Consequently, research on traffic safety remains a focal point in the fields of transportation and vehicle engineering. Particularly with the emergence of autonomous driving research, studies in traffic safety under autonomous or human-machine collaborative driving have become increasingly prominent.

Traffic safety and accident analysis cover a wide range of aspects. Broadly, they can be divided into static environment analysis and dynamic object state inspection in spatial terms, and pre-collision prediction and post-accident analysis in temporal terms. In static environment analysis, scholars have conducted comprehensive assessments and analyses of factors affecting traffic safety, including elements like pedestrian, vehicles, roads, and the environment [18, 5, 19]. These methods, although subjective, have been shown to reduce vehicle collisions and enhance traffic safety to some extent in redesigned and adjusted road environments, as evidenced by empirical studies. Additionally, in the area of dynamic object safety research, scholars have delved into modeling vehicle conflict and analyzing driver behavior, focusing on aspects like vehicle following relationships [20], driver compliance levels [21], and safety inspections in mixed environments at intersections [22].

Moreover, as accidents mark a causal boundary in traffic safety, pre-collision prediction and post-accident analysis are extremely important. Researchers have used various deep learning techniques, such as graph neural networks, to design traffic accident prediction networks [23, 24], aiming to predict traffic incidents. Some studies [12] analyze vehicle trajectories to assess the likelihood of collisions, striving to prevent them. Furthermore, analyzing accident data to identify causes and improve traffic safety is vital, whether through multivariate analysis of road traffic accident severity [25] or examining correlations between potential risk factors using data mining algorithms [26]. With the advent and evolution of autonomous vehicles, research exploring the factors influencing the severity of collisions involving autonomous cars [27] is emerging, seeking to assess the different impacts of autonomous and human driving patterns on collision severity for safety evaluation.

However, past research [23, 25, 26], whether based on comprehensive static environment analysis, dynamic vehicular kinematics or dynamics modeling, or through accident prediction and post-event evaluation, has been relatively independent. There lacks a holistic framework encompassing all these aspects, especially considering the advanced state of perception, prediction, and planning in autonomous driving. Proposing a complete end-to-end, comprehensive study of traffic safety is thus imperative.

### 2.2 Perception and planning

The mainstream approach in autonomous driving research has been to detect, recognize, and classify targets in the surrounding environment using sensors carried by vehicles or road-testing devices. This process serves the planning and control of vehicles. In the domain of target detection, research can be categorized into three main directions: single-vehicle intelligent perception, multi-vehicle collaborative perception, and vehicle-road collaborative perception.

In the realm of single-vehicle intelligent perception, landmark studies such as CenterNet [28] proposed a keypoint detection method, offering an efficient solution to explore visual patterns within individual cropped regions at minimal cost. ZoomNet [29] introduced a vehicle-centric camera-based target detection method with an adaptive zoom module, predicting higher-quality disparities from transformed recognition boxes. RTM3D [30] considered multiple perspective keypoints for 3D bounding boxes in image space, improving target detection rates in terms of speed and efficiency. FCOS3D [31] introduced a universal 3D target detection framework, decoupling 2D and 3D attributes from defined 7-DOF 3D target boxes, achieving top performance in visual-only 3D detection in the nuScenes dataset.

Recognizing the limitations of single-vehicle perception, some studies have focused on adopting a multi-vehicle collaborative approach to perception. CoBEVT [32] leveraged Vehicle-to-Vehicle (V2V) communication to share sensor information among vehicles, significantly improving perception performance and range, thereby collaboratively generating predictions of BEV maps. Coopernaut [33] similarly constructed a collaborative perception framework

using vehicle-to-vehicle communication for vision-based cooperative driving. It incorporated laser radar information, encoding it into a compact point-based representation for transmission between vehicles via real wireless channels. BEV-V2X [34] proposed an innovative approach based on V2X communication. It collected local BEV data from all connected and autonomous vehicles in the control area through roadside units or cloud centers, enabling the fusion and prediction of future global BEV occupancy grid maps. V2VFormer++ [35] built the first multimodal vehicle-to-vehicle collaborative perception framework, fusing individual camera-lidar representations with Dynamic Channel Fusion (DCF) in BEV space to capture global responses and achieve long-range collaborative perception.

Moreover, as road-testing devices capture higher angles and broader perspectives, covering specific scenes that individual vehicles may miss, some research has concentrated on integrating vehicle-road collaborative perception. To achieve Level 5 autonomous driving, the DAIR-V2X [36] dataset emphasized research on vehicle-infrastructure cooperation, releasing a large-scale, multimodal, multi-view vehicle-road collaborative dataset. BEVHeight [37] proposed a method for fusing vehicle-road collaborative perception, considering the rapid reduction in depth differences between vehicles and the ground as distance increases. It projected the height of road-testing cameras pixel-wise to depth, enhancing accuracy in occluded viewpoints and target detection under pose transformations. FecNet [38] fully considered static backgrounds captured by roadside lidar, presenting a feature-enhanced cascaded network. It improved the recognition ability of foreground features by extracting foreground information and fusing it with multi-stage feature maps.

In fact, past research on perception has mainly focused on the three mainstream directions of single-vehicle intelligent perception, multi-vehicle collaborative perception, and vehicle-road collaborative perception. However, research on integrating multi-vehicle collaborative perception with vehicle-road collaborative perception is still rare. If multiple devices with different perspectives and ranges can be further integrated, it will undoubtedly expand the range of perception and improve its accuracy.

### 2.3 LLM in transportation and intelligent vehicles

In the realm of LLM research, with ChatGPT as a prominent example, a new era in semantic parsing has been ushered in. Leading LLMs are generally based on transformer architectures for natural language processing, often equipped with tens of millions or even billions of parameters. These models are trained on extensive text datasets, enabling them to comprehend natural language and perform a wide range of complex tasks. The latest LLMs now possess dual capabilities in understanding both text and images. TransGPT introduces a versatile common-sense traffic model capable of predicting traffic situations, acting as an intelligent advisory assistant, supporting public transportation services, aiding in traffic planning and design, facilitating traffic safety education, assisting in management tasks, handling traffic accident reporting and analysis, and contributing to autonomous driving assistance systems.

The DiLu framework, as proposed by Wen et al. [16], combines reasoning and reflection modules, allowing the system to make decisions based on common-sense knowledge and evolve continuously. This infusion of knowledge-driven capabilities into autonomous driving systems, considering how humans drive, enables functionalities such as lane changing. TrafficGPT, as presented by Zhang et al. [17], integrates ChatGPT with foundational traffic models, enhancing the model’s ability to view, analyze, and process traffic data. This integration provides insightful decision support for managing urban traffic systems. Wang et al. [39] design a universal framework embedding LLM into the co-pilot role of vehicles. This framework enables the completion of specific driving tasks based on provided information, aligning with human intentions.

Previous work on LLM in the fields of traffic and autonomous driving has mainly focused on fine-tuning and training LLMs using traffic data. Alternatively, it involved secondary development of convenience applications based on LLM prompts or planning and control of vehicle movement based on LLM prompt words. However, the outstanding performance of LLM lies in its superior reasoning and judgment capabilities. Therefore, our in-depth integration and interaction of perception and trajectory prediction planning models with LLM aim to combine the perceptual work of computers with LLM research. This approach facilitates diverse tasks under human-machine interaction, including joint traffic situation warnings, accident responsibility allocation, accident causation analysis, and human-machine interaction prompts. This comprehensive integration effectively enhances the significance and value of perceptual work for both autonomous driving and human-machine collaborative driving, thereby advancing research and applications in traffic safety within the intelligent connected vehicle environment.

## 3 Methodology

### 3.1 Overview

The Overview of AccidentGPT, as depicted in Figures 1, presents a comprehensive structure emphasizing both the mission and the functional components of the system. The framework is designed to enhance traffic safety research,

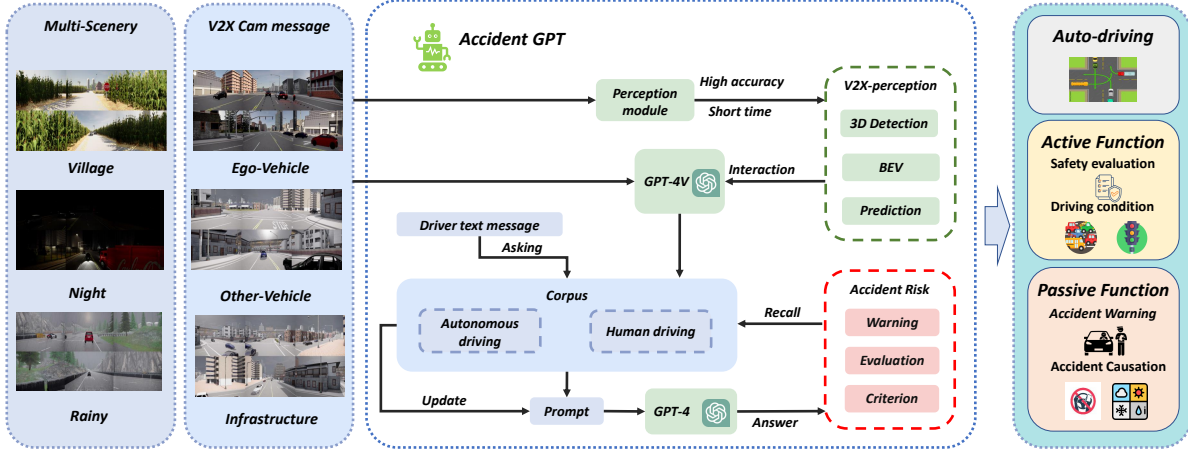


Figure 1: The overall framework of AccidentGPT.

including the human-machine interaction safety dialogue for artificially driven vehicles and traffic safety planning research for relevant departments.

AccidentGPT primarily consists of two main components. The first, V2X-perception, acts as the comprehensive scene perception module within the V2X architecture. It constructs an image extraction and analysis network using panoramic multi-camera information from multiple vehicles and roadside devices. Key elements of this module include the image processing backbone, view-transformer for image pose transformation, fusion-BEV for multi-scene fused perception, and a multi-head module for target detection, BEV perception, and trajectory prediction.

The second core component is the GPT-reasoning module, functioning as the reasoning interaction module under natural language. This includes the standard GPT-4V module, the multi-level priority sampling prompt module under corpus fusion, the active-passive task prompt module, and multiple output modules for accident-related tasks. This configuration aims to facilitate a seamless human-machine collaboration, leveraging machine language perception and human thought-driven knowledge to improve traffic safety.

To better elucidate our approach, we provide precise definitions for the proposed tasks as depicted in Figures 9. For the perception module, the task involves generating 3D object detection boxes, BEV perception, and trajectory prediction, utilizing inputs from multiple devices and perspectives. In terms of natural language reasoning tasks, the focus is on using the input and output information from the perception tasks as queried by the query module. This also includes constructing a learnable and updatable LLM corpus for prompts, enabling proactive prompts through drivers' active inquiries and passive prompts from the model, offering drivers passive cues such as advance notices and warnings for traffic safety events. Additionally, it involves responding to drivers' active inquiries and addressing tasks initiated by traffic authorities, like road traffic safety risk assessments and determining responsibilities in traffic accidents.

This comprehensive framework aims to address various tasks related to traffic and vehicle safety, particularly in the contexts of autonomous driving and human-machine interaction.

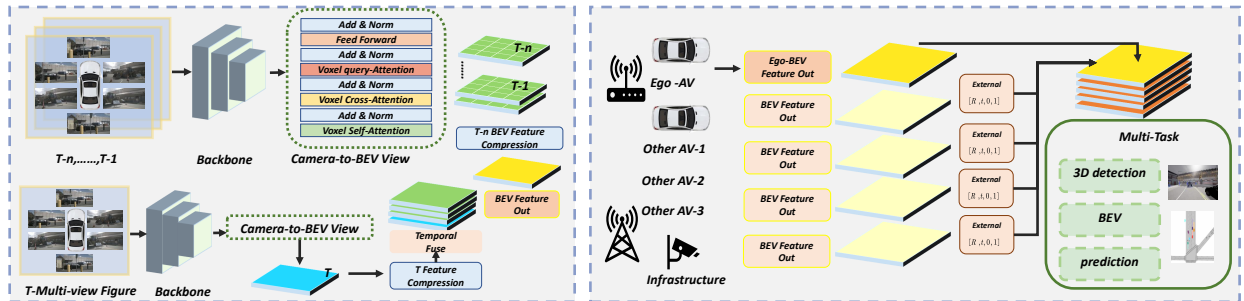


Figure 2: The V2X-perception architecture.



### 3.2 V2X-perception

We position the V2X-perception framework as a central component of AccidentGPT’s perceptual work. This framework seamlessly integrates and extracts features from the panoramic multi-images captured by multiple vehicles and road-testing devices. Through a well-constructed deep learning framework, it efficiently produces accurate 3D detection boxes for vehicles, perception results in the BEV perspective, and trajectory prediction results. These outputs serve as crucial inputs for GPT-reasoning, making the framework a vital element of AccidentGPT’s perceptual processes. The V2X-perception framework plays a pivotal role in enhancing the overall roadside traffic environment perception and map reconstruction. This will contribute to advancing collaborative safety driving for autonomous vehicles, long-range and blind-spot safety alerts for human-driven vehicles, human-machine interaction, and real-time assessments of traffic environments and accidents.

Specifically, in the realm of collaborative perception, drawing inspiration from BEVerse and DeepAccident, we have developed the global perception Module under the V2X architecture, referred to as V2X-perception as shown in Figure 2. For the input data from images, we applied preprocessing to each image based on the image-backbone to extract diverse information and features. Subsequently, the information captured by the surround-view cameras in each vehicle or device is input into the cross transformer network. Additionally, we established a Camera-to-BEV based on camera parameters, facilitating the transformation of information from various vehicles or road testing devices into a BEV perspective.

To enhance the comprehensiveness of global perception and recognition accuracy, we align the perception information extracted at different time sequences for each vehicle and road testing device under the transformation of Ego-motion-T. This alignment results in the construction of a BEV feature out that represents the complete state.

To optimize the output types, efficiency, and quality of tasks, we devised multiple detection heads to achieve various task outputs. These include the conventional 3D object detection, BEV perspective detection construction, and multi-vehicle trajectory prediction. These tasks uniformly employ BEV feature out for sampling and detection extraction, thereby enhancing the accuracy and completeness of multi-task perception work.

### 3.3 GPT-reasoning

The GPT-reasoning framework serves as the core element of AccidentGPT’s functionality. Through intricate fusion and extraction of perceptual information along with various collected data, we introduce multiple modules within the semantic mega-model of GPT-4V. These modules efficiently provide advanced proactive cues for human-machine collaboration and passive information retrieval. The goal is to enhance the semantic interaction framework for human-machine collaboration, addressing both proactive and reactive tasks in traffic safety warnings and alerts. This, in turn, aims to elevate the bidirectional safety mission between humans and computers.

Specifically, in the V2X-perception module, the constructed model achieves comprehensive collaborative perception for multiple vehicles and road testing, producing outputs such as 3D object detection and trajectory prediction. This module provides a priori knowledge for GPT-reasoning, facilitating more efficient and accurate perceptual information support.

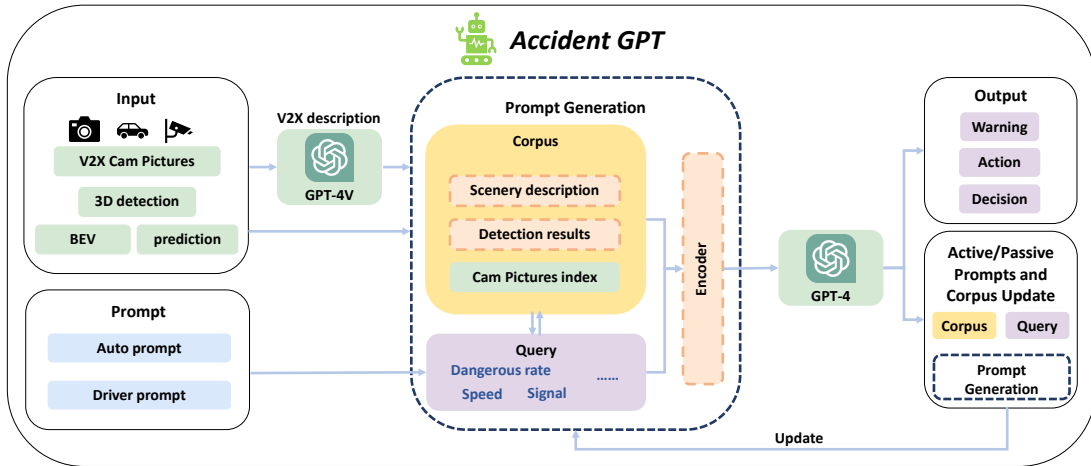


Figure 3: The GPT-reasoning module.

In the reasoning module, we define multiple decision tasks, establishing the multi-level priority sampling prompt module under corpus fusion, passive and active task prompt modules, and output modules for various accident-related tasks. This approach, supported by the standard GPT-4V framework, accurately perceives and describes decision tasks, offering advanced active and passive prompts to drivers or providing real-time road traffic environment assessments and accident causation analysis reports for traffic and planning departments. This helps reduce human-machine disparities, thereby enhancing human-vehicle interaction efficiency and ultimately reducing the occurrence of traffic accidents.

Specifically, the computational and interactive processes of our reasoning are illustrated in Figure 3, encompassing input, query, corpus sampling, prompt generation, LLM computation, active/passive prompts, corpus updates, and output modules. We will now elaborate on the design of these modules, emphasizing their positive impact on enhancing human-machine interaction efficiency and providing proactive and reactive safety prompts, as well as analysis reports, under perceptual fusion.

**Input Module:** The input serves as the header of the entire module, covering all relevant information inputs. Within the specific module inputs, there are primarily two parts: the input of perceptual information and the input of corpus information. For the perceptual information part, it requires inputs that encompass results from perception, including target detection, BEV, and trajectory prediction information. Additionally, it includes labeled raw image information. Regarding the input of corpus information, it mainly consists of language suggestions provided by the driver, the textual content of previous conversations, and suggested queries for the constructed corpus. All the aforementioned input information is not directly analyzed and understood by the LLM; instead, it is initially input into an information repository for further extraction and retrieval.

**Query Module:** This segment is a pivotal element in the system, offering proactive remote safety warnings and blind-spot alerts for human-driven vehicles. It can also provide comprehensive intelligent traffic safety analysis and reports for traffic management and planning departments based on prompts and suggestions. Specially, it enhances safety by providing improved driving recommendations and behavioral guidelines through human-machine dialogue and interaction. Drivers can set threshold levels for proactive warnings through dialogue, activating the proactive warning feature to ensure safe driving. When proactive prompts are enabled, there is continuous sampling and processing of image data from the vehicle and its surroundings. This image data serves as pre-input into the LLM during the query process. The results of image analysis are stored in a temporary corpus, enabling subsequent queries and retrievals. Once the triggered information reaches a specific intensity set by the threshold, it activates proactive warnings and prompts to ensure the driver's safe operations. In the absence of proactive prompts, the query module remains inactive and instead passively collects various information based on the driver's commands and suggestions. This includes both image and textual data, serving as pre-input into the LLM. The captured data is stored in a temporary corpus to support further processing, providing results and suggestions to the user side.

**Corpus sampling module:** After the processing by the aforementioned input and query modules, various types of information are efficiently transformed into the desired textual descriptions. This transformation facilitates the precise matching of descriptive content with the predefined and constructed corpus. In this section, a predefined corpus module is established to organize diverse and multimodal information into the "repository" of the corpus. Within this repository, various data is packaged and organized into standardized "boxes," accurately converting descriptive data into precise descriptive text. This further enables subsequent modules to select the optimal input information for LLM, supporting LLM in better understanding the current traffic situation and accelerating its reasoning speed.

**Prompt Generation Module:** This section, through the integration of preliminary work, establishes methods and suggestions for generating prompts that assist the large model in better understanding the input content and the objectives of the current task, thereby enhancing the quality of inference results. Specifically, the prompt generation module comprises three key components: the original information backup query retrieval module, the temporal corpus construction module for the current moment, and the corpus sampling module, which includes compressed and refined information from the corpus sampling module. In the original information backup query retrieval module, all raw image and text-type information from the current and historical data are provided. This supplementation aims to fulfill the specific needs of the LLM's inference and prompt tasks for the current task. The information encompasses historical images, trajectories, vectorized map data, and their corresponding specific labels. Additionally, the temporal corpus construction module dynamically builds a temporary corpus for short-term or user-defined interaction expectations at the current moment. This acts as a dynamic "cache" of information, facilitating rapid interaction, extraction, and analysis of environmental and driver prompt information for subsequent inference steps. The corpus sampling module serves as the primary information source for LLM computation. It includes standardized "boxes" representing various sources and modalities of information. The corpus sampling module integrates and filters information from these "boxes," extracting suitable prompt frameworks. The output from this module is then incorporated into the original information backup query retrieval module and the temporal corpus construction module, ultimately providing integrated and constructed input information for posing questions to the large model.

**LLM Computation Module:** In this phase, all prepared prompts are systematically input into the LLM to undergo intricate inference processes, expecting the LLM to make accurate decisions. Additionally, recognizing the inherent complexity and variability of driving scenarios and traffic safety environments, which may lead to unconventional biased decisions in semantic understanding, we adopt the chain-of-thought method to ensure the appropriate level of decision-making within the same input. This comprehensive approach integrates collaborative perception and reasoning to achieve improved human-vehicle interaction and contribute to advanced decision support systems for driving safety in intelligent connected vehicle environments. It also provides comprehensive recommendations for traffic management departments to monitor real-time road traffic safety risks.

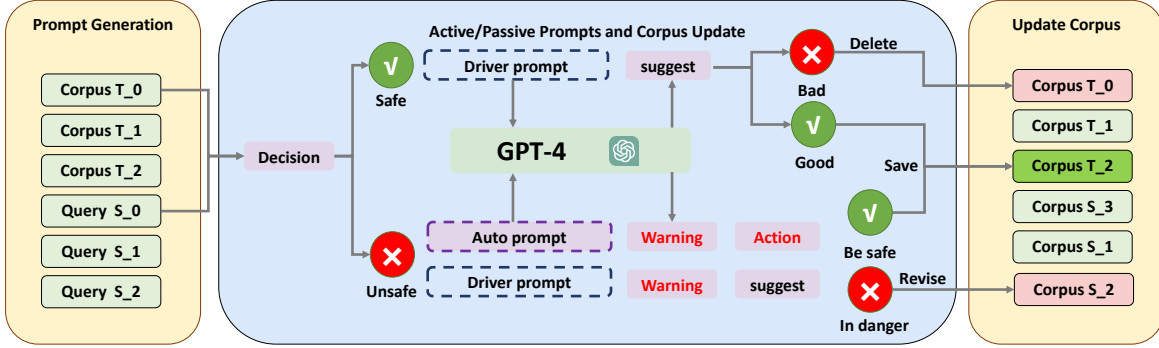


Figure 4: Active/Passive prompts and corpus update module.

**Active/Passive Prompts and Corpus Update Module:** In the process of providing proactive remote safety warnings and blind-spot alerts, as well as improved safety driving recommendations and behavioral guidelines through human-machine dialogue and interaction, we aim to implement two distinct sets of prompting rules for drivers, adjusting the prompt frequency to ensure safety margins. Specifically, within the active/passive prompts and corpus update module, we utilize the LLM to execute both proactive and reactive prompting tasks with the assistance of the proposed prompt module. The LLM evaluates whether the ongoing behavior aligns with the anticipated safety risk prompt range for the driver, enabling proactive prompts. The information prompted is then utilized as prior knowledge for subsequent prompts, accumulating valuable safety margin experience and enriching the corpus module. In situations where the system identifies significant safety risks, regardless of the driver’s prompt requests, risk prompts need to be issued, potentially involving adaptive driving strategies. For corpus updates, we introduce an iterative update method that continuously learns from past accident cases and decisions, gradually enhancing the module’s robustness and inference accuracy, elevating the inference capabilities to a higher dimension, as depicted in Figure 4. Specifically, during a driving task, we record the prompts used based on the input from the driving scenario and the corresponding decisions generated by the LLM for each decision frame. Once the driving session concludes without collisions or hazardous events, signifying a successful completion, AccidentGPT samples key decision frames from the sequence. These frames become integral for assessing risk experience, thereby enriching the corpus module. Conversely, if the current scenario leads to a dangerous situation, such as a collision due to misjudgments and decisions, indicating a system decision bias, it is semantically recorded to improve future decision outcomes. This iterative process gradually achieves the objective of enabling the AccidentGPT system to learn from its mistakes.

**Output Module:** Ultimately, the process involves extracting semantic content from the LLM representation and executing decisions and actions. The interconnection of each previously mentioned module culminates in the final decision, decoded from the feedback decoder in the LLM, thereby establishing a GPT-driven reasoning system. This comprehensive approach, incorporating both active and passive prompting, facilitates continuous learning through corpus updates and relies on LLM-driven decision-making.

This significantly contributes to the development of an advanced AccidentGPT system with adaptive learning capabilities, enabling effective interactive advice, tips, and traffic safety alerts to guide safety decisions in human-machine hybrid driving scenarios until fully autonomous driving is achieved. It also benefits traffic police and management agencies by enabling real-time human-machine interaction through the collaborative perception of global vehicles and road devices. This leads to intelligent analysis of the overall traffic safety environment and a comprehensive examination of the causes and responsibilities of accidents based on historical data from vehicle collisions.



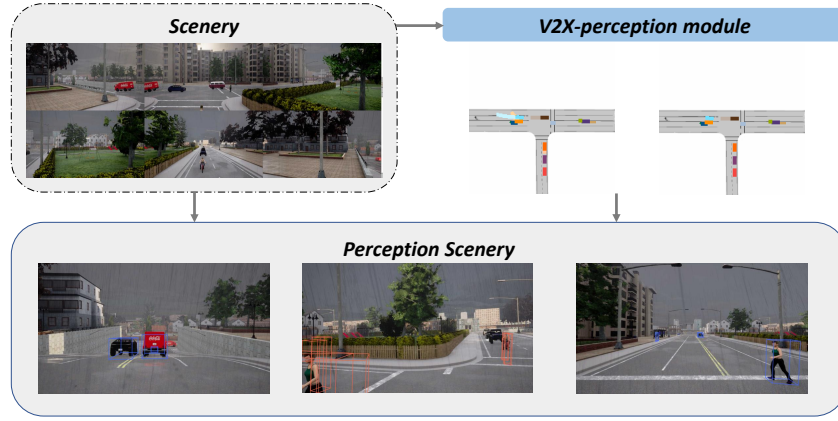


Figure 5: Automatic driving perception case study about AccidentGPT.

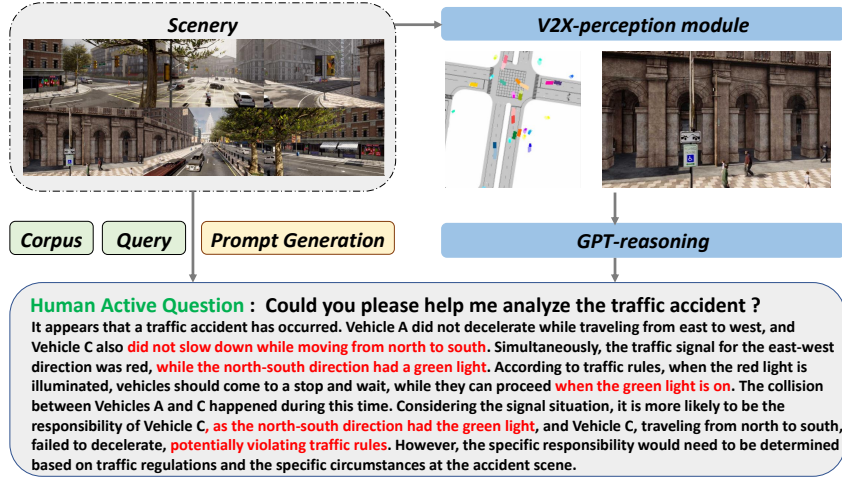


Figure 6: Active prompt case study about AccidentGPT.

## 4 Experimental results

### 4.1 Case study

To gain a deeper understanding of our work, this section provides three case study illustrating the primary functionalities of AccidentGPT.

In Figure 5, for autonomous vehicles, the model performs comprehensive environmental perception and understanding based on individual vehicle intelligence, vehicle-to-vehicle coordination, and vehicle-to-road coordination. It achieves precise scene perception, target detection, and trajectory prediction, enabling effective collision avoidance.

For human-driven vehicles, as shown in Figure 6, the driver actively inquires about the traffic situation. The AccidentGPT model utilizes the V2X perception module to interpret images, combines it with GPT reasoning network, and offers the driver judgments and predictions regarding vehicles, roads, and specific landmarks. The outcome of this interaction is providing warnings and suggestions to the driver. In Figure 7, guided by the safety alert threshold set by the driver in the previous dialogue, when the vehicle approaches a complex intersection or area, the large model takes a proactive stance, providing safety driving recommendations. It suggests the driver to exercise caution at signal-less intersections and emphasizes the visibility challenges posed by backlit conditions during dusk. This proactive approach aims to issue timely warnings to the driver based on the current environmental conditions. Therefore, it can provide proactive

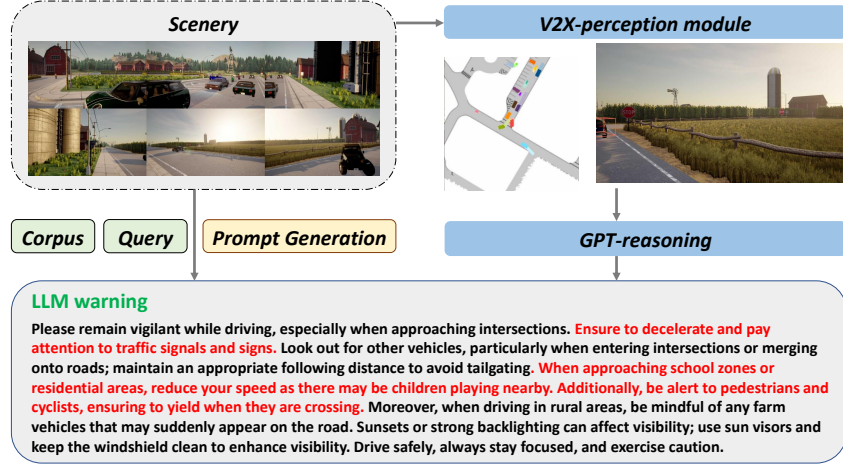


Figure 7: Passive prompt case study about AccidentGPT.

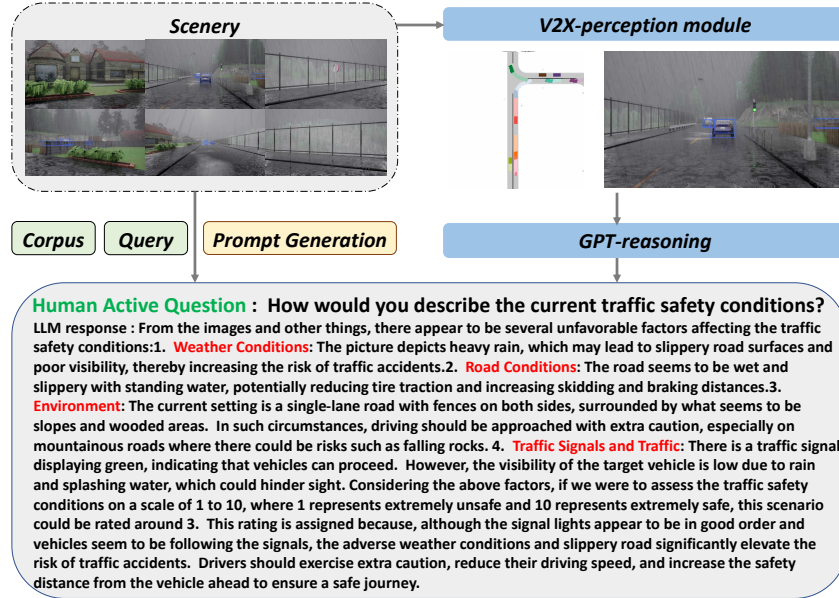


Figure 8: Traffic environment analysis case study about AccidentGPT.

long-range safety warnings and blind-spot alerts to the driver and offer better safety driving recommendations and behavioral guidelines through human-machine dialogue and interaction.

In Figure 8, for traffic police and management agencies, intelligent and real-time traffic safety analysis and comprehensive evaluation reports regarding pedestrian, vehicles, roads, and the environment can be achieved through collaborative perception from multiple vehicles and road testing devices. This includes in-depth analysis of traffic safety, providing a holistic assessment and analysis of accident causes and responsibilities following vehicle collisions.

## 4.2 Settings

In the experimental section, we will conduct tests and experiments on the proposed model to demonstrate its effectiveness and superiority. For perception and planning tasks, we will perform tests based on the DeepAccident dataset. This dataset includes perception results from six devices, comprising 4 cars' surround-view camera information and two road testing devices. The environment includes collisions with illegal activities such as running red lights and unexpected

**Main function about AccidentGPT**

Active									
INPUT	Ego Cam	BEV	trajectory	weather	V2X Cam	vehicle	Detection	Road	OUTPUT
Safety evaluation	✓	✗	✗	✓	✓	✗	✓	✓	Safety level/suggestion
Driving condition	✓	✓	✓	✗	✗	✗	✗	✗	Safety level/suggestion
Traffic condition	✓	✓	✗	✗	✓	✓	✗	✓	Condition/suggestion
Traffic violation	✓	✗	✓	✗	✗	✓	✗	✗	Condition/suggestion
Passive									
Accident prediction	✓	✓	✗	✗	✓	✗	✓	✗	Safety suggestion
Accident responsibility	✓	✓	✓	✗	✓	✗	✗	✗	Division of responsibility
Causation analysis	✗	✗	✓	✓	✓	✗	✓	✗	Analysis /suggestion
Traffic situation	✓	✓	✗	✗	✓	✓	✗	✗	Dangerous level

Figure 9: Main function about AccidentGPT.

Table 1: Comparison of detection and motion results.

Different model	Index		
	Detection (%)	Motion	
		mIOU (%)	VPQ (%)
Average Fusion [12]	36.2	52.1	39.5
DiscoNet [40]	38.5	54.2	42
V2X-ViT [41]	40.1	55.1	43.2
CoBEVT [32]	40.8	56.2	44
V2X-perception (our)	<b>41.07</b>	<b>57.3</b>	<b>45.2</b>

collisions during driving. Additionally, it contains various types of intersections and weather conditions for selection and evaluation. For testing based on LLM, we utilized OpenAI’s GPT-4 for semantic tasks restructuring and integration, and GPT-4V for continuous frame image recognition and analysis. In our constructed AccidentGPT framework, we also established a corpus module for updating and learning to facilitate semantic and encoding transformations. In the experimental part, we will test and validate V2X-perception from aspects like image perception and prediction to demonstrate its superiority in 3D object detection, BEV perception, and trajectory prediction. For the GPT-reasoning module, we will evaluate its performance on eight different mainstream tasks, as shown in Figure 9. We will conduct separate assessments and comparisons with benchmark methods and mainstream classification models. Ablation experiments will also be carried out to demonstrate the beneficial value and effectiveness of each proposed module.

### 4.3 Perception and planning

**Perception and Trajectory Prediction Performance.** As shown in Table 1, we conducted tests and validations by comparing our proposed V2X-perception model with other mainstream collaborative models, namely DiscoNet[40], V2X-ViT [41], and CoBEVT [32]. It is evident that the V2X-perception model outperforms previous modules in both 3D target detection and trajectory prediction tasks, as judged by the three evaluation metrics. It consistently achieved the best performance, thereby providing more beneficial guidance for the subsequent GPT-reasoning module in terms of perception.

**Performance Based on Different Types of Vehicle Visibility.** During observation, the visibility of different types of vehicles or pedestrians is a crucial aspect of perception. Therefore, we simultaneously assessed the perceptual model’s visibility differences for various vehicle or pedestrian types, as shown in Table 2. Specifically, we calculated the mean Average Precision (mAP) for each category, which measures the average accuracy; mean Average Surface Error (mASE), assessing the average error between the detected target surface and the actual surface; mean Average Orientation Error (mAOE), evaluating the average error between the detected target orientation and the actual orientation; and mean Average Velocity Error (mAVE), measuring the average error between the predicted and actual velocities. The results

Table 2: Comparison of different type results.

Test model type	Metric			
	mATE (m)	mASE (1-IoU)	mAOE (rad)	mAVE (m/s)
Car	0.994540292	0.964026	1.022296	1.267359
Truck	0.995231552	0.957066	1.01609	1.205831
Van	0.997434757	0.986803	1.012704	1.103993
Pedestrian	1.001661385	0.999283	1.025322	1.026851

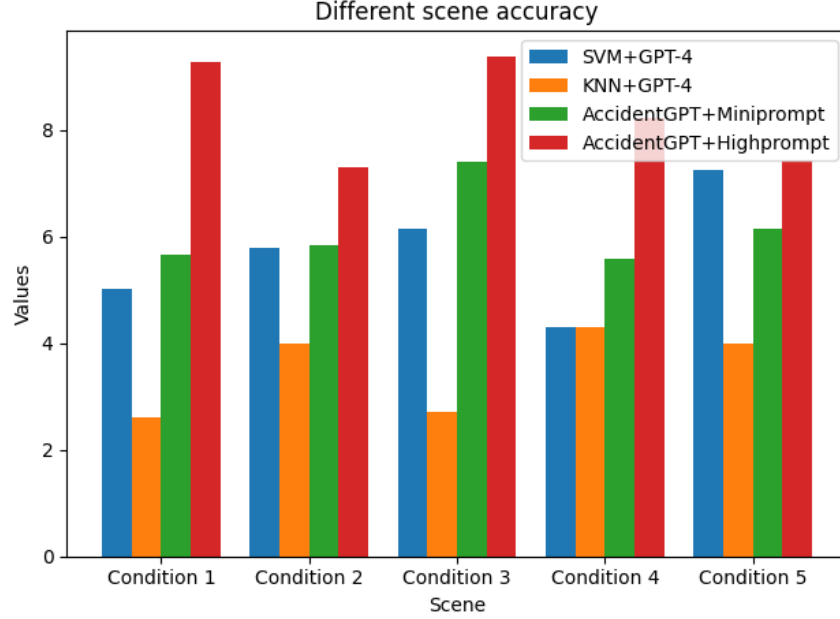


Figure 10: Comparison of partition Models in different scenarios' outcomes.

indicate that our model (V2X-perception) performs well in the perception of various vehicles and pedestrians, achieving a commendable level of effectiveness.

#### 4.4 GPT-reasoning

**Model performance in defining and classifying events.** To assess the performance of event definition, we compared our proposed end-to-end AccidentGPT method with the classification efficiency of SVM [42], KNN [43] pre-processing combined with AccidentGPT. We conducted the analysis using four evaluation metrics:

- Accuracy: The proportion of correctly classified samples among all samples.
- Precision: The proportion of true positive samples among those classified as a particular category.
- Recall: The proportion of correctly classified samples among all samples belonging to a particular category.
- F1 Score: The harmonic mean of precision and recall.

Regardless of the scenario definition, traffic risk delineation, or even accident responsibility assignment, the results obtained by conventional deep learning models were consistently lower than our proposed end-to-end AccidentGPT method, as shown in Table 3. We selected a specific scenario task and compared the SVM+AccidentGPT and KNN+AccidentGPT methods. These methods did not effectively judge and analyze images, leading to significant errors in subsequent semantic analysis. In contrast, the end-to-end AccidentGPT method more effectively captured accurate information, enabling precise judgment and decision-making, as illustrated in Figure 10.

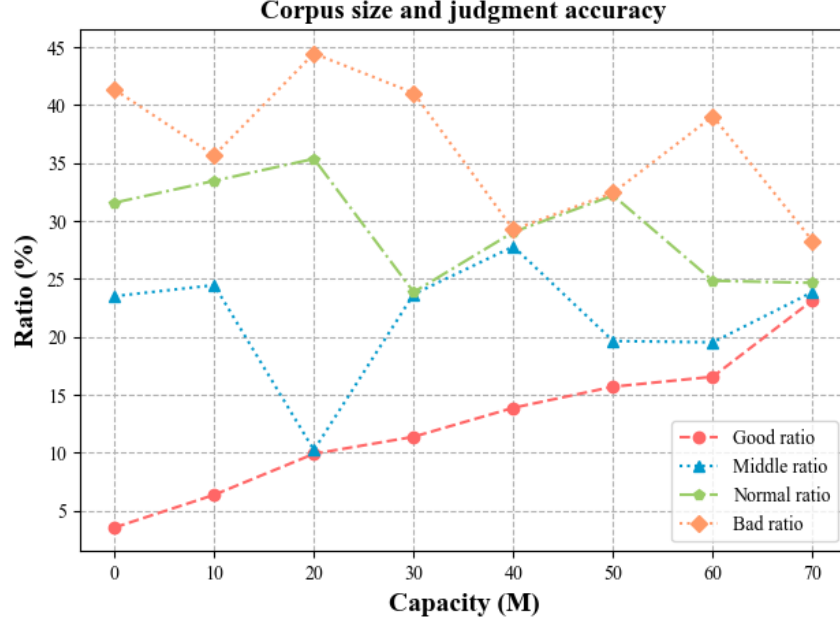


Figure 11: Comparison of AccidentGPT models in different corpus size.

Table 3: Comparison of different partition model results

Partition model	Type	Index			
		Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM+GPT-4 [42]	Scene	0.53	0.42	0.34	0.34
	Safety	0.74	0.78	0.78	0.74
	Accident	0.74	0.51	0.5	0.49
KNN+GPT-4 [43]	Scene	0.47	0.29	0.3	0.26
	Safety	0.16	0.15	0.23	0.16
	Accident	0.47	0.3	0.3	0.27
AccidentGPT (GPT-reasoning+GPT-4)	Scene	<b>0.89</b>	<b>0.83</b>	<b>0.83</b>	<b>0.84</b>
	Safety	<b>0.89</b>	<b>0.92</b>	<b>0.89</b>	<b>0.89</b>
	Accident	<b>0.89</b>	<b>0.92</b>	<b>0.83</b>	<b>0.85</b>

**Performance of Prompt Generation Intensity in Different Scenarios.** The intensity of prompt generation can significantly influence the accuracy of the model’s output. We conducted experiments to analyze the impact of prompts with different intensities, categorizing the content volume of prompt generation as (Mini/Middle/High+Prompt), as illustrated in Table 4. It is evident that as the prompt generation intensity gradually increases, the performance of prompt results also improves across different scenarios.

**Performance of Model’s Prompt Library Update Frequency Ratio on Model Inference.** The effectiveness of the model is, to a certain extent, attributed to the proposed prompt library update module. When the library is gradually updated based on feedback, it effectively provides deeper and more accurate guiding prompts for further semantic information. As indicated in Table 5, our analysis of prompt library update frequency reveals that as the library undergoes in-depth updates along with prompts, the accuracy of prompts in different corresponding scenarios will



Table 4: Comparison of AccidentGPT models in different prompt.

Different prompt	Misson	Index				
		Number	Good (%)	Middle (%)	Normal (%)	Bad (%)
Mini+prompt	Safety evaluation	1	2.61	26.18	32.08	39.12
	Driving condition	2	5.62	13.08	26.24	55.06
	Traffic condition	3	3.81	7.76	35.79	52.64
	Traffic violation	4	4.12	28.19	32.65	35.04
	Accident prediction	5	3.33	21.59	31.78	43.3
	Accident responsibility	6	3.99	10.77	41.2	44.04
	Causation analysis	7	5.2	21.58	34.37	38.85
	Traffic situation	8	4.52	5.2	43.39	46.89
Middle+prompt	Safety evaluation	1	7.43	16.21	32.6	43.76
	Driving condition	2	7.85	23.57	30.32	38.26
	Traffic condition	3	7.76	10.97	39.06	42.2
	Traffic violation	4	10.24	19.46	25.79	44.51
	Accident prediction	5	10.87	27.05	27.18	34.89
	Accident responsibility	6	8.32	25.34	25.42	40.92
	Causation analysis	7	8.46	19.35	36.05	36.14
	Traffic situation	8	7.14	22.15	33.87	36.84
High+prompt	Safety evaluation	1	10.83	16.34	31.87	40.96
	Driving condition	2	13.03	15.46	31.74	39.76
	Traffic condition	3	16.41	19.54	31.72	32.33
	Traffic violation	4	15.72	20.92	27.31	36.05
	Accident prediction	5	9.42	29.17	29.83	31.58
	Accident responsibility	6	17.56	25.08	25.97	31.39
	Causation analysis	7	16.62	19.03	32.08	32.26
	Traffic situation	8	13.12	17.8	21.18	47.9

progressively improve. This improvement is particularly remarkable in the "good" category. In Figure 11, the variation in prompt update capacity further underscores this point.

## 5 Conclusion

In order to establish a unified framework for a comprehensive understanding and application of traffic safety, aiming to enhance road traffic safety during the transition from manual to fully autonomous driving, this paper introduces an end-to-end accident analysis and prevention from V2X environmental perception with multi-modal large model AccidentGPT. This framework achieves three-dimensional object detection, BEV perception, and vehicle trajectory prediction across the entire road environment by constructing a comprehensive perception model under the V2X architecture. The integration of a natural language reasoning module with perception methods is designed to blend cognitive reasoning abilities with perceptual approaches. This integration is aimed at addressing various challenges in accident analysis and prevention within the field of traffic safety. Extensive experimental validation demonstrates that each proposed module exhibits significant potential and robust generalization capabilities in perception and safety tasks. Furthermore, AccidentGPT seamlessly integrates with various typical autonomous driving models and datasets, emphasizing its deployment potential in human-machine hybrid driving scenarios and traffic safety optimization and governance systems. To the best of our knowledge, we are the first to research and develop a large-scale model that combines perception with human cognition, integrating comprehensive scene perception and human understanding into

Table 5: Comparison of AccidentGPT models in different prompt renewal rate.

Renewal rate	Misson	Index				
		Number	Good (%)	Middle (%)	Normal (%)	Bad (%)
10% Renewal prompt	Safety evaluation	1	2.87	22.81	31.7	42.62
	Driving condition	2	4.82	20.49	35.98	38.71
	Traffic condition	3	3.56	23.52	31.58	41.34
	Traffic violation	4	6.46	13.33	22.01	58.19
	Accident prediction	5	6.07	11.15	41.36	41.42
	Accident responsibility	6	4.76	8.63	32.04	54.57
	Causation analysis	7	3.3	3.97	34.28	58.44
	Traffic situation	8	2.53	12.63	41.06	43.79
30% Renewal prompt	Safety evaluation	1	9.05	25.79	26.08	39.08
	Driving condition	2	2.38	14.89	40.1	42.63
	Traffic condition	3	4.34	24.9	31.7	39.06
	Traffic violation	4	6.38	24.47	33.47	35.67
	Accident prediction	5	9.73	21.03	22.6	46.64
	Accident responsibility	6	6.77	19.08	19.93	54.21
	Causation analysis	7	6.49	27.49	29.9	36.11
	Traffic situation	8	9.92	10.24	35.37	44.46
50% Renewal prompt	Safety evaluation	1	12.46	16.07	26.15	45.31
	Driving condition	2	11.39	23.69	23.87	41.05
	Traffic condition	3	14.1	25.88	27.31	32.71
	Traffic violation	4	19.45	21.17	24.69	34.69
	Accident prediction	5	15.87	16.03	33.25	34.86
	Accident responsibility	6	16.47	24.68	29.09	29.76
	Causation analysis	7	13.89	27.8	29.05	29.26
	Traffic situation	8	11.16	26.63	30.06	32.15
70% Renewal prompt	Safety evaluation	1	13.68	19.77	19.94	46.61
	Driving condition	2	10.47	25.15	26.76	37.61
	Traffic condition	3	16.31	25.23	28.7	29.75
	Traffic violation	4	16.57	19.54	24.85	39.04
	Accident prediction	5	11.16	13.96	20.28	54.6
	Accident responsibility	6	18.35	22.34	24.96	34.34
	Causation analysis	7	15.71	19.65	32.21	32.44
	Traffic situation	8	11.44	15.11	19.53	53.93
90% Renewal prompt	Safety evaluation	1	23.17	23.88	24.67	28.28
	Driving condition	2	20.76	22.5	26.28	30.47
	Traffic condition	3	20.26	22.29	28.51	28.95
	Traffic violation	4	20.34	22.14	26.22	31.31
	Accident prediction	5	20.34	21.81	26.53	31.32
	Accident responsibility	6	15.95	16.77	22.82	44.46
	Causation analysis	7	14.85	27.03	28.63	29.49
	Traffic situation	8	14.77	22.44	27.25	35.54

traffic safety research. The demonstrated capabilities highlight its long-term application potential from current manual driving scenarios to the complete transition to autonomous driving systems, aiming to improve efficiency and ultimately reduce the occurrence of traffic accidents.

## Data Availability

Data page: <https://deepaccident.github.io>

## References

- [1] World Health Organization. Health topics, <https://www.who.int/health-topics/>.
- [2] Yuping Hu, Ye Li, Helai Huang, Jaeyoung Lee, Chen Yuan, and Guoqing Zou. A high-resolution trajectory data driven method for real-time evaluation of traffic safety. *Accident Analysis & Prevention*, 165:106503, 2022.
- [3] Yilong Ren, Han Jiang, Nan Ji, and Haiyang Yu. Tbsm: A traffic burst-sensitive model for short-term prediction under special events. *Knowledge-Based Systems*, 240:108120, 2022.
- [4] Rui An, Zhaomin Tong, Yimei Ding, Bo Tan, Zihao Wu, Qiangqiang Xiong, and Yaolin Liu. Examining non-linear built environment effects on injurious traffic collisions: A gradient boosting decision tree analysis. *Journal of Transport & Health*, 24:101296, 2022.
- [5] Mehrnaz Asadi, Mehmet Baran Ulak, Karst T Geurs, Wendy Weijermars, and Paul Schepers. A comprehensive analysis of the relationships between the built environment and traffic safety in the dutch urban areas. *Accident analysis & prevention*, 172:106683, 2022.
- [6] Le-Ning Wang. Real-time evaluation method of vehicle conflict risk on urban expressway based on smartphone gps data. *Journal of advanced transportation*, 2021:1–13, 2021.
- [7] Yasir Ali, Anshuman Sharma, Md Mazharul Haque, Zuduo Zheng, and Mohammad Saifuzzaman. The impact of the connected environment on driving behavior and safety: A driving simulator study. *Accident Analysis & Prevention*, 144:105643, 2020.
- [8] Lianzhen Wang, Yuping Wang, Lingyun Shi, and Huizhi Xu. Analysis of risky driving behaviors among bus drivers in china: The role of enterprise management, external environment and attitudes towards traffic safety. *Accident Analysis & Prevention*, 168:106589, 2022.
- [9] Wanbao Ye, Chuanlin Wang, Fuxiang Chen, Shuzhen Yan, and Liping Li. Approaching autonomous driving with cautious optimism: analysis of road traffic injuries involving autonomous vehicles based on field test data. *Injury prevention*, 27(1):42–47, 2021.
- [10] Megan S Ryerson, Jordan E Miller, and Flaura K Winston. Edge conditions and crash-avoidance roles: the future of traffic safety in the world of autonomous vehicles, 2019.
- [11] Shahzeb Ansari, Fazel Naghdy, and Haiping Du. Human-machine shared driving: Challenges and future directions. *IEEE Transactions on Intelligent Vehicles*, 7(3):499–519, 2022.
- [12] Tianqi Wang, Sukmin Kim, Wenxuan Ji, Enze Xie, Chongjian Ge, Junsong Chen, Zhenguo Li, and Ping Luo. Deepaccident: A motion and accident prediction benchmark for v2x autonomous driving. *arXiv preprint arXiv:2304.01168*, 2023.
- [13] Yushan Han, Hui Zhang, Huifang Li, Yi Jin, Congyan Lang, and Yidong Li. Collaborative perception in autonomous driving: Methods, datasets and challenges. *arXiv preprint arXiv:2301.06262*, 2023.
- [14] Yue Hu, Shaoheng Fang, Zixing Lei, Yiqi Zhong, and Siheng Chen. Where2comm: Communication-efficient collaborative perception via spatial confidence maps. *Advances in neural information processing systems*, 35:4874–4886, 2022.
- [15] R OpenAI. Gpt-4 technical report. arxiv 2303.08774. *View in Article*, 2:3, 2023.
- [16] 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. *arXiv preprint arXiv:2309.16292*, 2023.
- [17] Siyao Zhang, Daocheng Fu, Zhao Zhang, Bin Yu, and Pinlong Cai. Trafficgpt: Viewing, processing and interacting with traffic foundation models. *arXiv preprint arXiv:2309.06719*, 2023.

- [18] Marta Rodrigues Obelheiro, Alan Ricardo da Silva, Christine Tessele Nodari, Helena Beatriz Bettella Cybis, and Luis Antonio Lindau. A new zone system to analyze the spatial relationships between the built environment and traffic safety. *Journal of transport geography*, 84:102699, 2020.
- [19] Aleksandar Stevanovic, Jelka Stevanovic, Jaehyun So, and Marija Ostojic. Multi-criteria optimization of traffic signals: Mobility, safety, and environment. *Transportation Research Part C: Emerging Technologies*, 55:46–68, 2015.
- [20] Fangfang Zheng, Can Liu, Xiaobo Liu, Saif Eddin Jabari, and Liang Lu. Analyzing the impact of automated vehicles on uncertainty and stability of the mixed traffic flow. *Transportation research part C: emerging technologies*, 112:203–219, 2020.
- [21] Anshuman Sharma, Zuduo Zheng, Jiwon Kim, Ashish Bhaskar, and Md Mazharul Haque. Assessing traffic disturbance, efficiency, and safety of the mixed traffic flow of connected vehicles and traditional vehicles by considering human factors. *Transportation research part C: emerging technologies*, 124:102934, 2021.
- [22] Ramin Arvin, Asad J Khattak, Mohsen Kamrani, and Jackeline Rio-Torres. Safety evaluation of connected and automated vehicles in mixed traffic with conventional vehicles at intersections. *Journal of Intelligent Transportation Systems*, 25(2):170–187, 2020.
- [23] Le Yu, Bowen Du, Xiao Hu, Leilei Sun, Liangzhe Han, and Weifeng Lv. Deep spatio-temporal graph convolutional network for traffic accident prediction. *Neurocomputing*, 423:135–147, 2021.
- [24] Da-Jie Lin, Mu-Yen Chen, Hsiu-Sen Chiang, and Pradip Kumar Sharma. Intelligent traffic accident prediction model for internet of vehicles with deep learning approach. *IEEE transactions on intelligent transportation systems*, 23(3):2340–2349, 2021.
- [25] Deyu Wang, Qinyi Liu, Liang Ma, Yijing Zhang, and Haozhe Cong. Road traffic accident severity analysis: A census-based study in china. *Journal of safety research*, 70:135–147, 2019.
- [26] Sharaf Alkheder, Fahad AlRukaibi, and Ahmad Aiash. Risk analysis of traffic accidents’ severities: An application of three data mining models. *ISA transactions*, 106:213–220, 2020.
- [27] Weixi Ren, Bo Yu, Yuren Chen, and Kun Gao. Divergent effects of factors on crash severity under autonomous and conventional driving modes using a hierarchical bayesian approach. *International journal of environmental research and public health*, 19(18):11358, 2022.
- [28] Kaiwen Duan, Song Bai, Lingxi Xie, Honggang Qi, Qingming Huang, and Qi Tian. Centernet: Keypoint triplets for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6569–6578, 2019.
- [29] Zhenbo Xu, Wei Zhang, Xiaoqing Ye, Xiao Tan, Wei Yang, Shilei Wen, Errui Ding, Ajin Meng, and Liusheng Huang. Zoomnet: Part-aware adaptive zooming neural network for 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12557–12564, 2020.
- [30] Peixuan Li, Huaici Zhao, Pengfei Liu, and Feidao Cao. Rtm3d: Real-time monocular 3d detection from object keypoints for autonomous driving. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 644–660. Springer, 2020.
- [31] Tai Wang, Xinge Zhu, Jiangmiao Pang, and Dahua Lin. Fcos3d: Fully convolutional one-stage monocular 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 913–922, 2021.
- [32] Runsheng Xu, Zhengzhong Tu, Hao Xiang, Wei Shao, Bolei Zhou, and Jiaqi Ma. Cobevt: Cooperative bird’s eye view semantic segmentation with sparse transformers. *arXiv preprint arXiv:2207.02202*, 2022.
- [33] Jiaxun Cui, Hang Qiu, Dian Chen, Peter Stone, and Yuke Zhu. Coopernaut: End-to-end driving with cooperative perception for networked vehicles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17252–17262, June 2022.
- [34] Cheng Chang, Jiawei Zhang, Kunpeng Zhang, Wenqin Zhong, Xinyu Peng, Shen Li, and Li Li. Bev-v2x: Cooperative birds-eye-view fusion and grid occupancy prediction via v2x-based data sharing. *IEEE Transactions on Intelligent Vehicles*, pages 1–18, 2023.
- [35] Hongbo Yin, Daxin Tian, Chunmian Lin, Xuting Duan, Jianshan Zhou, Dezong Zhao, and Dongpu Cao. V2vformer ++: Multi-modal vehicle-to-vehicle cooperative perception via global-local transformer. *IEEE Transactions on Intelligent Transportation Systems*, 2023.
- [36] Haibao Yu, Yizhen Luo, Mao Shu, Yiyi Huo, Zebang Yang, Yifeng Shi, Zhenglong Guo, Hanyu Li, Xing Hu, Jirui Yuan, et al. Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21361–21370, 2022.

- [37] Lei Yang, Kaicheng Yu, Tao Tang, Jun Li, Kun Yuan, Li Wang, Xinyu Zhang, and Peng Chen. Bevheight: A robust framework for vision-based roadside 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21611–21620, 2023.
- [38] Ziren Gong, Zhangyu Wang, Guizhen Yu, Wentao Liu, Songyue Yang, and Bin Zhou. Fecnet: A feature enhancement and cascade network for object detection using roadside lidar. *IEEE Sensors Journal*, 23(19):23780–23791, 2023.
- [39] Shiyi Wang, Yuxuan Zhu, Zhiheng Li, Yutong Wang, Li Li, and Zhengbing He. Chatgpt as your vehicle co-pilot: An initial attempt. *IEEE Transactions on Intelligent Vehicles*, 2023.
- [40] Eloi Mehr, Ariane Jourdan, Nicolas Thome, Matthieu Cord, and Vincent Guitteny. Disconet: Shapes learning on disconnected manifolds for 3d editing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3474–3483, 2019.
- [41] Runsheng Xu, Hao Xiang, Zhengzhong Tu, Xin Xia, Ming-Hsuan Yang, and Jiaqi Ma. V2x-vit: Vehicle-to-everything cooperative perception with vision transformer. In *European conference on computer vision*, pages 107–124. Springer, 2022.
- [42] Thorsten Joachims. Making large-scale svm learning practical. Technical report, Technical report, 1998.
- [43] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. Knn model-based approach in classification. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, pages 986–996. Springer, 2003.



## Appendix

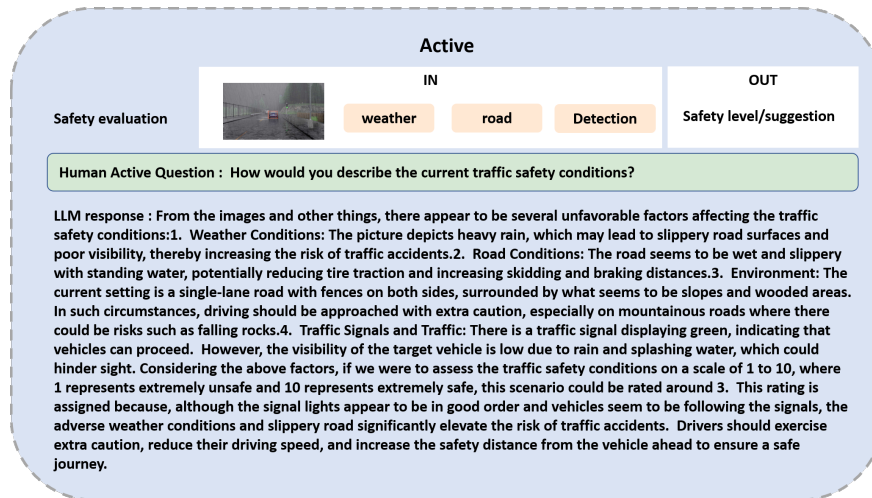


Figure 12: Sample of safety evaluation in AccidentGPT.

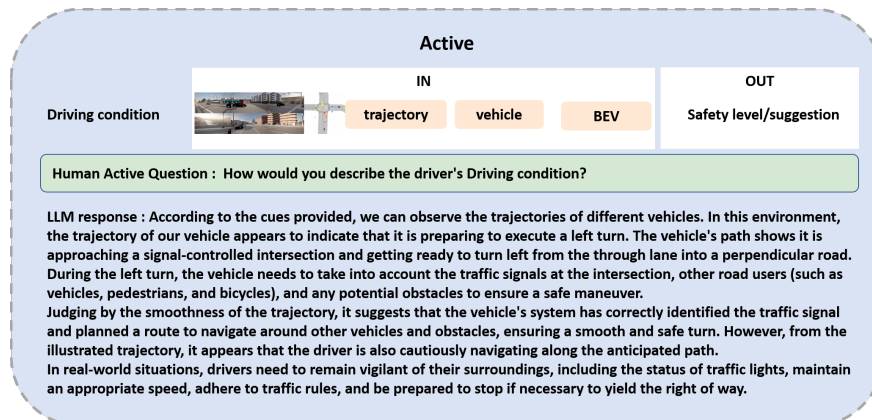


Figure 13: Sample of driving condition in AccidentGPT.


Active			
	IN		OUT
Traffic condition		V2X Cam Road	Condition/suggestion
<p>Human Active Question : Could you please help me analyze the current traffic situation?</p> <p>LLM response : From the movement of vehicles from different directions and the visible road markings, this appears to be a four-way intersection. In the center of the intersection, two vehicles are moving towards each other; a blue vehicle is proceeding straight from one side of the intersection, while a red vehicle is making a left turn, and it seems their paths might intersect. Typically, the right of way in such situations is determined by traffic rules. The red vehicle making a left turn should yield to the oncoming traffic, unless it has a green left-turn signal. Note that there is a "STOP" marking on the ground, indicating that vehicles should come to a complete stop before proceeding. Vehicles adhering to this marking usually wait for oncoming traffic to pass or respect the signals.</p> <p>In terms of traffic flow, aside from these two vehicles moving towards each other, the overall traffic on the road doesn't seem too busy. In the distance, some vehicles can be seen entering or preparing to enter the intersection from other exits. While the overall environment appears calm, the possibility of a collision between the two vehicles exists, considering there are no traffic signals for proper control, or drivers might not be adhering to traffic rules.</p> <p>It's crucial to emphasize that drivers should always remain vigilant, follow local traffic regulations, and be observant to ensure the safe passage through the intersection, particularly in the absence of proper traffic control signals.</p>			

Figure 14: Sample of traffic condition in AccidentGPT.


Active			
	IN		OUT
Traffic violation		Ego Cam vehicle	Condition/suggestion
<p>Human Active Question : Please tell me about potential traffic violations that could occur?</p> <p>LLM response : Based on a comprehensive assessment of the images and other available information, there doesn't appear to be any evidence of traffic violations at the moment. However, this assessment relies on detailed knowledge of local traffic regulations. Nevertheless, we can make speculative observations about some potential infractions that are visible:</p> <ol style="list-style-type: none"> <li>1. Running a Red Light: If a vehicle fails to come to a stop when the traffic light is red, it constitutes a violation. In the left foreground, there is a car at the intersection seemingly stopped at a red light. Whether it fully stopped before the red light or not cannot be ascertained from the image, determining compliance with traffic rules.</li> <li>2. Wrong-Way Driving or Illegal Lane Change: Driving against the flow of traffic on a one-way road or making unauthorized lane changes is illegal. Although the image doesn't overtly show wrong-way driving or illegal lane changes, such activities beyond the visible areas would be considered violations.</li> <li>3. Pedestrian Crossing: If pedestrians cross the road without a crosswalk or during a red light, it could be deemed illegal. A pedestrian is seen crossing in the middle in front of the vehicle; if it was during a green light and within a crosswalk, it would be legal.</li> <li>4. Emergency Vehicles: A firetruck is seen crossing the intersection in the middle on the left. Emergency vehicles often have exemptions from regular traffic signals during missions but must ensure public safety isn't compromised.</li> </ol> <p>Traffic regulations vary between countries and regions, so determining whether an action constitutes a violation depends on local laws.</p>			

Figure 15: Sample of traffic violation in AccidentGPT.


Passive			
	IN		OUT
Accident prediction		BEV V2X Cam Detection	Safety suggestion
<p>LLM warning : Please remain vigilant while driving, especially when approaching intersections. Ensure to decelerate and pay attention to traffic signals and signs. Look out for other vehicles, particularly when entering intersections or merging onto roads; maintain an appropriate following distance to avoid tailgating. When approaching school zones or residential areas, reduce your speed as there may be children playing nearby. Additionally, be alert to pedestrians and cyclists, ensuring to yield when they are crossing. Moreover, when driving in rural areas, be mindful of any farm vehicles that may suddenly appear on the road. Sunsets or strong backlighting can affect visibility; use sun visors and keep the windshield clean to enhance visibility. Drive safely, always stay focused, and exercise caution.</p>			

Figure 16: Sample of accident prediction in AccidentGPT.

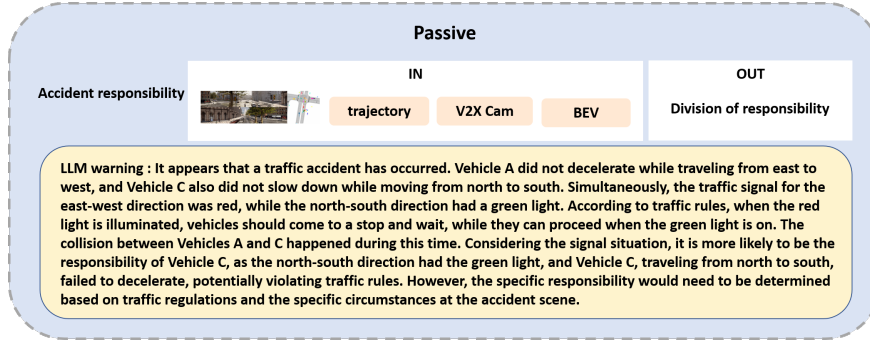


Figure 17: Sample of accident responsibility in AccidentGPT.

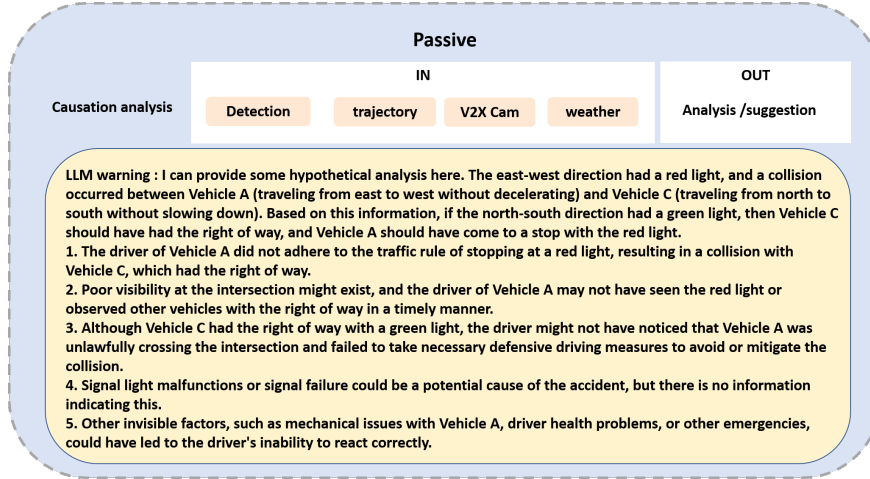


Figure 18: Sample of causation analysis in AccidentGPT.

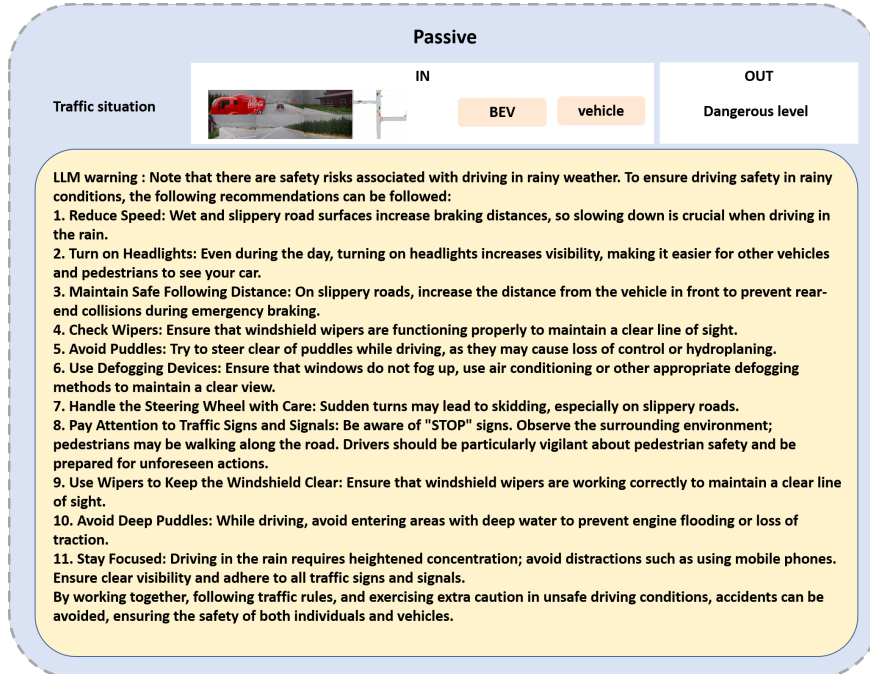


Figure 19: Sample of traffic situation in AccidentGPT.