

Article

Occupant-Aware Decision-Making with Large Vision-Language Model for Autonomous Vehicles

Titong Jiang , Xinyu Zhao , Xuewu Ji  and Yahui Liu * 

School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China

* Correspondence: liuyahui@tsinghua.edu.cn

Abstract

Autonomous driving (AD) has emerged as a transformative technology that holds the potential to free humans from the need for manual driving and provide a safer, more comfortable and efficient driving experience. However, most AD systems make decisions solely based on vehicle dynamics and environmental factors such as road conditions and surrounding vehicles, while the occupant's mental states, such as subjective feelings and experience, are neglected. As a result, autonomous vehicles (AVs) often fail to meet the occupant's physical and mental demands, ultimately leading to a compromised driving experience. In this study, we propose an occupant-aware decision-making paradigm (ODP) for AD systems. ODP first perceives the occupant's physical and physiological states that are closely related to mental states, such as facial expressions and physiological signals, through the occupant monitoring system (OMS). Then, a large vision-language model (VLM) processes the occupant's physical and physiological states via the chain of thought (CoT) technique to analyze the occupant's mental states and infer the occupant's needs. Finally, the VLM makes driving decisions that match the occupant's demands and preferences. Experimental results show that ODP can make decisions that are significantly better aligned with the occupant's actual needs than existing methods.

Keywords: autonomous driving; decision-making; vision-language-action model; large language model; chain-of-thought

1. Introduction

Autonomous driving (AD) has been widely considered as a revolutionary technology that promises to improve the safety, comfort, efficiency, and convenience of road transportation [1]. In recent years, AD systems have received substantial attention from both academia and industry and made significant progress, enabling autonomous vehicles (AVs) to navigate and make decisions in real-time without human intervention. However, despite the fact that the functionality of AVs has been improving over the years, the general public's view on AVs remains conservative, as the annual American Auto Association (AAA) polls show that the acceptance rate of people towards self-driving cars has not increased for many years [2]. In other words, there is a significant mismatch between people's perceived usefulness and risk and the actual capabilities of AVs, which hinders the future development and deployment of AD technology [3].

Previous studies [4,5] have shown that this mismatch stems from the AD systems' inability to understand the occupant's needs and preferences. Take driving style as an example, different people prefer different driving styles. Some people prefer a more conservative driving style, while others prefer a more aggressive driving style.



Academic Editor: Ossama Mokhiamar

Received: 23 January 2026

Revised: 18 February 2026

Accepted: 22 February 2026

Published: 25 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Moreover, the same person may switch preferences under different circumstances, e.g., a more aggressive driving style when running late for an important meeting, and a more conservative driving style when driving with children on board. Most AD systems, with only vehicle dynamics and environmental factors like road conditions and surrounding vehicles as inputs, cannot adjust their driving styles dynamically and only provide a few predefined driving styles as options. With the occupant's subjective feelings and experience absent, AVs fail to meet the occupant's physical and mental demands in real-time, ultimately leading to a compromised driving experience.

Currently, personalized decision-making in autonomous driving systems can be generally divided into three categories: explicit mode selection, implicit adaptation from historical data, and environment-centric decision-making [6,7]. Explicit mode selection is the process of explicitly selecting a driving mode based on the driver's preferences, such as sport mode or comfort mode. While straightforward, this simple approach cannot adapt to the dynamic mental states of the occupant or contextual changes of the environment. Implicit adaptation from historical data refers to the process of learning from the driver's past behavior and preferences to adapt the driving mode accordingly. Typical implicit adaptation methods include inverse reinforcement learning (IRL) and continual learning. This approach assumes that preferences are stable over time, and thus ignores the transient emotional states of the occupant. Environment-centric decision-making considers the environmental factors (e.g., traffic conditions, weather, and traffic participants) when making driving decisions. In this approach, the occupant is excluded from the feedback loop, leading to suboptimal travel experiences. While all of these aforementioned approaches have been extensively explored in the existing literature, they unanimously assume that the occupant's preferences and demands are static. None of them have the ability to understand and meet the occupant's real-time experiences and needs. To develop AVs that can actually understand and meet the occupant's needs, it is of great importance to incorporate the occupant's mental states into the decision-making process of AD systems [8,9].

In order to achieve this goal, three main challenges need to be overcome: (1) how to measure the occupant's mental states including subjective feelings and experience, (2) how to model the occupant's mental activities and obtain the occupant's needs from the occupant's mental states, and (3) how to make driving decisions that match the occupant's demands and preferences.

For the first challenge, it is not straightforward to measure the occupant's mental states, as they are subjective and dynamic. In practice, physical and physiological signals, such as facial expressions, body gestures, and heart rate signals, have been widely used in psychological and ergonomical studies as proxies to measure human's mental states [10]. For example, one's facial expression can reflect their subjective feelings and experience, while a high heart rate can indicate that the person is under stress. Fortunately, for AVs, the occupant monitoring system (OMS), which is equipped with advanced functions including facial expression recognition and remote physiological monitoring, has been popular in recent years. The prevalence of OMS enables us to measure the occupant's physical and physiological signals, thereby inferring the occupant's mental states.

For the second challenge, traditional methods for mental activity analysis are limited in their ability to capture the complex and dynamic nature of the occupant's mental states. For example, rule-based methods [11] and machine learning methods [12] that rely on handcrafted features are limited in their ability to handle the vast amount of data generated by OMS. Recent advances in large language models (LLMs) have demonstrated their strong capabilities in understanding and reasoning about mental activities [13]. In addition, the introduction of the chain of thought (CoT) technique, which allows LLMs to generate

a series of intermediate reasoning steps verbally before giving the final answer, further enhances the LLMs' reasoning capabilities and stimulates better interpretability. Therefore, AD systems can analyze and understand the occupant's mental activities and actual needs by leveraging LLMs together with the CoT technique.

For the third challenge, it is essential to develop a decision-making paradigm that can match the occupant's demands and preferences. Large vision-language models (VLMs), which incorporates visual modality into LLMs, enable LLMs to directly perceive the world through images and videos and understand the visual content in a more natural and intuitive way. Previous studies [14] have shown that VLMs, after proper fine-tuning, are capable of safely navigating through complex and dynamic environments. However, existing models only focus on the safety and efficiency aspect of driving, whereas the occupant's actual needs are neglected. Also while large-scale AD data are available, these data do not include occupant information, and hence cannot be used to train VLMs to understand the occupant's needs. As a result, specialized AD data that include occupant information as well as the occupant's needs and preferences are in great demand to adapt AD systems for occupants.

In this study, we propose an occupant-aware decision-making paradigm (ODP) for AD systems. As discussed above, ODP consists of three stages: states monitoring, mental analysis, and decision-making. In the states monitoring stage, we use OMS to monitor the occupant's physical and physiological signals, such as facial expressions, body gestures, and heart rate signals. The occupant's physical and physiological signals, together with AD perception results, are then sent to the mental analysis stage, which leverages a VLM and the CoT technique to analyze the occupant's mental activities and infer the occupant's needs from the occupant's mental states. In the decision-making stage, the VLM makes driving decisions that match the occupant's demands based on the analysis from the second stage. Moreover, a large-scale occupant-centric decision-making dataset is carefully curated to support the development and validation of ODP. The main contributions of this study can be summarized as follows:

- We propose an occupant-aware decision-making paradigm for AD systems, which is capable of perceiving, analyzing, and reasoning about the occupant's mental activities and needs, and making driving decisions that match the occupant's demands and preferences.
- We develop a large-scale occupant-centric decision-making dataset, which not only includes naturalistic driving data, but also includes the occupant's states, subjective feelings, and needs.
- We evaluate the performance of ODP on the developed dataset, and show that its decision-making capabilities can effectively match the occupant's demands and preferences.

1.1. Related Work

Recent advances in AD have largely focused on improving perception, prediction, decision-making, and planning modules using deep learning and large-scale datasets. However, the integration of human-centered factors, particularly the physical and mental demands of occupants, into the decision-making pipeline remains underexplored. This section reviews related work across three key areas: (1) occupant monitoring system in AVs, (2) occupant-aware AD, and (3) VLMs in AD.

1.1.1. Occupant Monitoring System

The ability to detect the driver's or passenger's cognitive and affective states has long been studied in the context of an advanced driver assistance system (ADAS) [15]. The primary goal of OMS is to determine if the driver or passenger is in a safe or unsafe

state, e.g., distraction [16], drunk driving [17], drowsiness [18], or fatigue [19]. As such, OMS is usually equipped with facial cameras with facial expression recognition, eye gaze detection, and head pose detection functions. Recently, with the rapid advances of sensor technology, remote physiological measurement, which can directly measure the occupant's vital physiological signals including heart rate and breathing rate, is also available for occupant impairment monitoring in OMS [20]. However, to the best of our knowledge, this information has not been used to analyze the occupant's mental activities.

1.1.2. Mental State Analysis

Mental state analysis is a critical step in understanding the occupant's needs and preferences. Traditionally, mental state analysis has been implemented in the human–vehicle interface (HVI) module of the intelligent cockpits for human–vehicle interaction [21]. For example, emotion recognition approaches have been widely implemented in in-vehicle infotainment systems to facilitate natural and effective communication between the occupant and the vehicle [22]. Mental state analysis has also been used in driver monitoring systems for distraction [23], fatigue [24], and drowsiness [25] detection. However, these systems are designed for human–vehicle interaction rather than decision-making, and thus do not analyze the occupant's mental activities to infer the occupant's needs. In contrast, large language models have demonstrated their strong capabilities in understanding and reasoning about mental activities [26,27], making them a promising tool for analyzing the occupant's mental states in AD systems. Unfortunately, to the best of our knowledge, previous studies have not explored the use of LLMs for mental state analysis in AD systems, and thus the potential of LLMs in this domain remains untapped.

1.1.3. Occupant-Aware Autonomous Driving

A growing body of work recognizes that an one-size-fits-all driving policy is insufficient to satisfy diverse user preferences [28]. Some studies [29] have explored personalization through explicit user input, e.g., driving style selection. Others attempt implicit adaptation by learning individual preferences from historical driving data using inverse reinforcement learning (IRL) [30] or continual learning [31]. However, these methods typically assume static preferences regardless of driving context or the occupant's real-time demands, resulting in suboptimal results. In another streamline of research, contact-based physiological sensors, such as Electrocardiogram (ECG) [32], electromyograph (EMG) [33], and electroencephalograph (EEG) [34], have been used to understand the occupant's subjective feelings and experience. However, these studies are only suitable for laboratory settings as it is impractical to deploy contact-based sensors in commercial vehicles.

1.1.4. Large Vision-Language Models in Autonomous Driving

VLMs, such as LLaVa [35], GPT-4V [36], and Qwen-VL [37], have demonstrated remarkable zero-shot reasoning capabilities in solving open-world tasks. In the field of AD, VLMs have attracted increasing scholarly attention due to their exceptional ability to understand complex visual scenes, identify key obstacles, make feasible decisions, and handle corner cases [38–40]. In one of the pioneering works, DriveVLM [39] proposes to employ a VLM for scene understanding and decision-making, generating meta-actions for further planning and control. However, these systems remain environment-centric and do not consider the internal state of the vehicle's occupant.

2. Materials and Methods

In this section, we describe the proposed ODP paradigm and the dataset curated for this study in detail.

2.1. Occupant-Aware Decision-Making Paradigm

The proposed ODP paradigm consists of three stages, states monitoring, mental analysis, and decision-making, as illustrated in Figure 1.

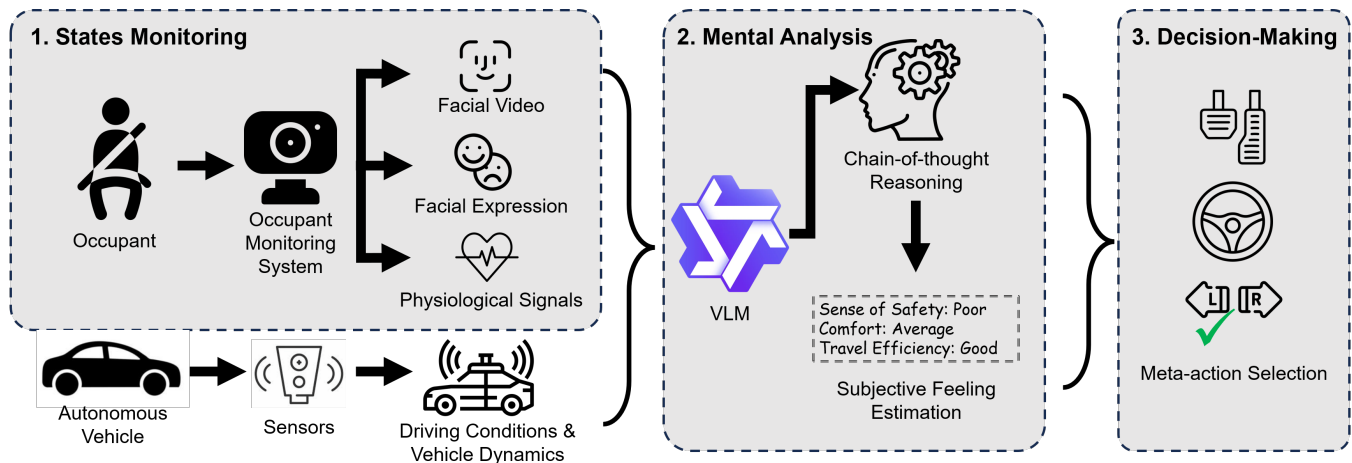


Figure 1. Overview of the proposed occupant-aware decision-making paradigm (ODP). ODP consists of three stages: states monitoring, mental analysis, and decision-making.

In the states monitoring stage, we leverage OMS to monitor the occupant's physical and physiological signals. For physical signals, the facial camera of OMS is used to capture the occupant's facial video, and facial expression recognition technology is applied for occupant emotion detection. For physiological signals, we use remote physiological measurement enabled by facial camera to monitor the occupant's heart rate (HR) and heart rate variability (HRV). The Baevsky stress index (BSI) [41] is then calculated from HRV to represent the stress level of the occupant. Please note that remote physiological measurement is prone to noise and artifacts from the driving scenarios. Therefore, robust remote physiological measurement plays an important role in the overall success of the ODP paradigm. Readers are referred to [15,20,41] for more information on how the occupant's states can be reliably monitored in AVs.

From the states monitoring stage, the occupant's states, including facial video, emotion, HR, HRV, and stress level, are monitored and recorded. In the mental analysis stage, a VLM is used to analyze the occupant's states and infer the occupant's needs and preferences. Here, we use Qwen3-VL-8B-Instruct [42] after careful consideration. First, the Qwen3-VL series models are one of the most advanced open-source VLMs, which has demonstrated outstanding performance in solving real-world high-level tasks. Second, the model size of the VLM is a key factor in developing VLM-based AD systems: a large model is infeasible to be deployed on AVs with limited computing resources, while a small model may lead to inferior performance. Previous studies [40] have shown that VLMs with a moderate model size, e.g., from 5B to 10B, are suitable for AD tasks. Therefore, this study chooses the 8B variant of the Qwen3-VL series as the backbone. Third, while the VLM has already been trained on a large-scale and diverse dataset, enabling broad general knowledge and reasoning capabilities, it still requires further specialized fine-tuning for occupant-aware decision-making tasks. Therefore, the Qwen3-VL-8B-Instruct version is chosen over the Qwen3-VL-8B-Thinking version as the former is more flexible and recommended for fine-tuning.

In the mental analysis stage, the VLM is responsible for systematically analyzing the occupant's states, the vehicle states, and the driving conditions, and ultimately understanding the occupant's needs. This task involves complicated reasoning across multiple modalities and domains, rendering it an extremely challenging problem for the VLM. To boost the reasoning capability of the VLM, we employ the CoT technique by dividing the whole task into three subtasks: scene understanding, mental activity analysis, and subjective feeling estimation. In the scene understanding subtask, the VLM is prompted to analyze the vehicle states and the driving conditions, and determine the key elements that may affect driving decisions, e.g., surrounding vehicles, road conditions, traffic lights, and weather conditions. In the mental activity analysis subtask, the VLM is prompted to process the occupant's states from the states monitoring stage, and infer the occupant's mental activities such as stress, fear, and relaxed. In the subjective feeling estimation subtask, we divide the occupant's subjective feelings into three categories, including sense of safety, comfort, and travel efficiency, and prompt the VLM to estimate the occupant's subjective feelings based on the analysis from the previous two subtasks. In this way, the occupant's needs, such as the need for a safer, more comfortable or more efficient driving experience, can be inferred from the occupant's subjective feelings. Through the meticulously designed CoT prompts, the VLM is able to not only accurately understand the occupant's needs via in-depth reasoning, but also generate verbal analysis that is interpretable and reliable, which is essential for building trustworthy AD systems.

In the decision-making stage, the VLM makes driving decisions that match the occupant's needs based on the analysis from the mental analysis stage. Here, we follow DriveVLM [39] and define 16 meta-actions that represent the basic driving actions, including but not limited to acceleration, deceleration, stopping, lane changing, turning left and right, and minor lateral and longitudinal adjustments. A comprehensive list of meta-actions can be found in Section 2.3. The VLM is prompted to select 1 meta-action from the whole set that best suits the analysis from the previous stage. Subsequently, a dedicated planning and control (PnC) module is employed to translate the meta-actions into trajectories and control signals.

2.2. Dataset

To support the development and validation of ODP, we curate a large-scale occupant-centric decision-making dataset that includes naturalistic driving data as well as the occupant's states, subjective feelings, and needs. The dataset is collected from 42 Chinese participants who are unfamiliar with AVs, representing potential buyers, including age groups from 18 to 61 years old and gender ratios of 52% male and 48% female. Among all participants, 14 have no driving experience. All participants are in good health without any known neurological or psychiatric disorders, well informed about the purpose of this study, and provided written informed consent before the experiment. Each participant is paid 100 CNY after the experiment. The experiments are conducted on a Geely Xingyue L SUV and an IM LS6 SUV respectively, with Geely driven by a professional driver and IM driven by the onboard navigation on autopilot (NOA) AD system and supervised by a professional driver. In total, 24 participants took only the Geely test, 10 participants took the IM test, and 8 participants took both tests. The test route, as shown in Figure 2, is approximately 16 km long and encompasses diverse scenarios including urban roads, highways, intersections, roundabouts, tunnels, and overpasses. Figure 3 shows the experimental apparatus of the experiments, which consists of two Logitech C920e webcams, a Contec CMS50E pulse oximeter, and a laptop. The participant is seated at the front passenger seat, with a webcam mounted on the dashboard to capture the participant's facial video, and the pulse oximeter is clipped to the participant's fingertip to

measure the participant's HR and HRV. The occupant's facial expression and BSI are then derived from the facial video and HRV. Please note that while HR and HRV are measured using a pulse oximeter during experiments, in ODP, HR and HRV are directly calculated from the occupant's facial videos using remote photoplethysmography algorithms [20]. Moreover, another front-facing Logitech C920e webcam is mounted on the windshield to capture the driving scene, and the vehicle dynamics are extracted from the CAN BUS signals of the test vehicles. All of the data above are collected, synchronized, and recorded by the laptop in real-time using a custom software.

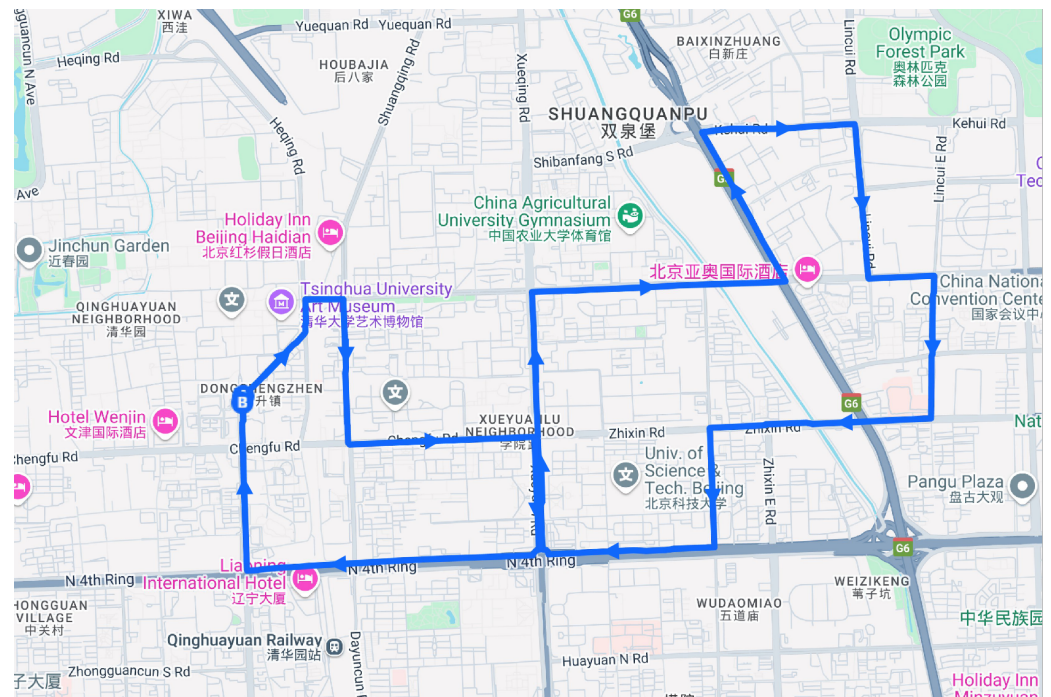


Figure 2. The test route of this study. The test route, located in Beijing, China, is approximately 16 km long and encompasses diverse scenarios including urban roads, highways, intersections, roundabouts, tunnels, and overpasses.

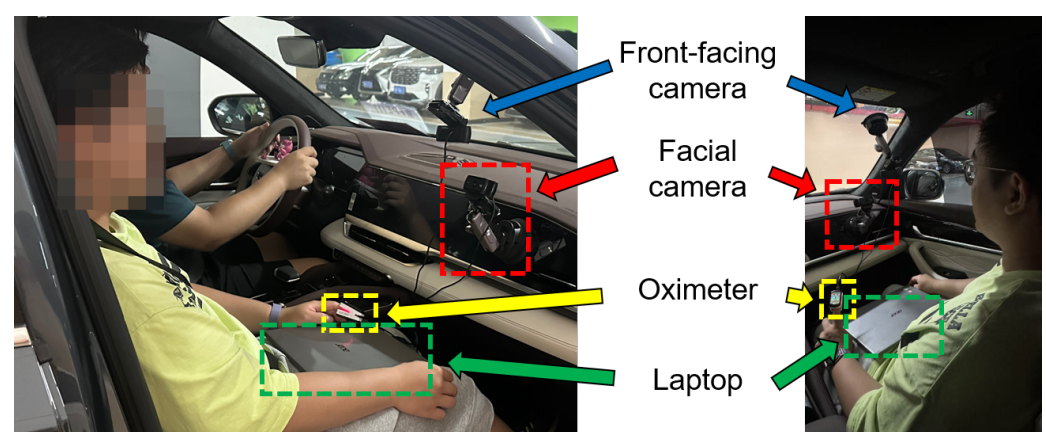


Figure 3. The experimental apparatus of this study. The front-facing camera captures the driving scene, the facial camera records the occupant's facial video, the pulse oximeter measures the occupant's BVP, and the laptop is used to collect and synchronize data.

Aside from the automatically collected data, we also collect participant-annotated data, including the participant's subjective feelings and driving demands. For the subjective feelings, we ask the participant to rate their sense of safety, comfort, and travel efficiency on a 5-point Likert scale at an interval of every 2 min, which is recorded by the experimenter

seated in the back seat. To ensure that the reported ratings are consistent among participants, instead of a numerical rating, we ask participants to rate subjective feelings verbally using descriptions including very good, good, medium, poor, and very poor. Moreover, we explicitly ask participants to not avoid using extreme descriptions in order to reduce central tendency bias. For the driving demands, we develop a mobile phone app on which each meta-action is a button, and instruct the participant to press corresponding buttons to record their desired driving decisions regardless of the vehicle's actual actions. For example, if the vehicle is following the preceding car too closely that the participant feels unsafe, the participant should press the *Follow from a greater distance* button, even though the vehicle's actual action is to follow closely. Before conducting the experiments, we ask the participants to install the app on their own mobile phones and familiarize themselves with the operations of the app. Ideally, participants should record desired driving decisions in each scene. However, in pilot tests, we find that asking the participant to record every desired decision is very exhausting, and that the vehicle's actual actions can often match the participant's desired decisions. As such, we ask the participant to record desired driving decisions only when the vehicle's actual actions do not match the participant's needs, e.g., when the actual actions make the participant feel unsafe, uncomfortable, or inefficient. In other cases, we manually label the desired driving decisions according to the vehicle's actual actions. During manual labeling, an experimenter examines the recorded videos and records the vehicle's actual actions. Meanwhile, Qwen3-VL-Plus is prompted to verify if the recorded actions of the experimenter are correct. In this way, manual labeling errors can be detected and corrected. All data are carefully scrutinized after the experiments by the experimenters.

In summary, the dataset includes 31.4 h of high quality occupant-aware driving data with 15,251 decision-making annotations. The dataset is randomly split into a training set of 12,201 annotations and a test set of 3050 annotations. Please note that the data split is participant-agnostic, i.e., the same participant may be present in both sets. For each decision-making annotation, we extract the following features from the last 5 s of the annotating time to form a sample: facial video and facial expression labels from the facial webcam, HR, HRV and BSI from the pulse oximeter, vehicle dynamics from the CAN BUS signals, driving scenario video from the front-facing camera, and the participant's subjective feelings and desired driving decisions.

2.3. Model Training

The VLM is fine-tuned on the training set of the developed dataset using supervised fine-tuning (SFT). The input to the VLM includes the occupant's facial video, facial expression label, HR, HRV, BSI, vehicle dynamics, and driving scenario video, formulated as the following prompt:

You are a senior expert in autonomous driving decision-making analysis.

Task

Based on the following input information, please analyze the driving scenario the autonomous vehicle is currently in and infer the occupant's current states.

Input

1. Vehicle Dynamics:

The vehicle's current speed is v km/h, longitudinal acceleration is a_x m/s², lateral acceleration is a_y m/s², and angular velocity is w_z deg/s.

2. Driving Condition Video:

V_{drive}

3. Occupant's Facial Video:

V_{face}

4. Occupant's States:

The occupant's current heart rate, heart rate variability, and Baevsky stress index are x_{HR} bpm, x_{HRV} ms, and x_{BSI} , respectively. The facial expression is x_{exp} .

Output

Based on the above information, please analyze the current driving scenario of the autonomous vehicle and the occupant's state, and infer the occupant's subjective feelings and driving decisions through a clear and logical chain of thought. The specific content includes:

1. Describe the key visual elements in the driving scenario (such as road type, traffic participants, traffic light status, obstacles, lane lines, weather/lighting conditions, etc.), and determine the vehicle's current behavior.
2. Analyze the occupant's facial video and physiological states to infer the occupant's emotional state (e.g., nervous, relaxed, confused, surprised, etc.).
3. Analyze the occupant's subjective feelings and driving decisions.

Requirements

1. Subjective feelings are divided into three dimensions: sense of safety, comfort, and travel efficiency. Each dimension is divided into five levels: very good, good, medium, poor, and very poor.
2. Driving decisions should be selected from the following options: remain constant speed, move forward slowly, left turn, right turn, acceleration, deceleration, rapid acceleration, rapid deceleration, lane change to the left, lane change to the right, slight left adjustment, slight right adjustment, follow at a greater distance, follow at a closer distance, stop and wait, and overtake.
3. Output should be in JSON format, and the JSON output should contain five key-value pairs: "chain of thought", "sense of safety", "comfort", "travel efficiency", and "driving decision".

Here, v is the vehicle's current speed, a_x is the longitudinal acceleration, a_y is the lateral acceleration, w_z is the angular velocity, V_{drive} is the driving condition video captured by the front facing camera, V_{face} is the occupant's facial video, x_{HR} is the occupant's current heart rate, x_{HRV} is the occupant's heart rate variability, x_{BSI} is the occupant's Baevsky stress index, and x_{exp} is the occupant's current facial expression. Thus, the input contains all the necessary information for the VLM to make occupant-aware driving decisions. It is noted that high-resolution videos may incur high computational costs for VLMs. Consequently, the resolution of the 5 s input videos are first downsampled to 640×360 before being fed into the VLM. As for the frame rate, Qwen3-VL has an internal automatic frame extraction mechanism; thus, the frame rate is automatically determined by the VLM.

The output of the VLM includes three parts: CoT for mental analysis, subjective feeling estimation, and driving decision. The CoT part is a step-by-step reasoning process that guides the VLM to analyze the occupant's mental activities before estimating subjective feelings and making decisions. To help the VLM produce valid and effective reasoning CoTs, we propose a CoT synthesis method that generates high-quality CoT samples automatically from the training set. Specifically, we leverage Qwen3-VL-Plus, one of the most advanced VLMs available, to generate initial CoT samples using the following prompt:

You are a senior expert in autonomous driving decision-making analysis.

Task

Based on the following input information, please analyze the current driving scenario of the autonomous vehicle, infer the occupant's current state, and generate a clear and logically rigorous chain of thought from **input** to **results**.

Input

1. Vehicle Dynamics:

The vehicle's current speed is v km/h, longitudinal acceleration is a_x m/s², lateral acceleration is a_y m/s², and angular velocity is w_z deg/s.

2. Driving Condition Video:

V_{drive}

3. Occupant's Facial Video:

V_{face}

4. Occupant's States:

The occupant's current heart rate, heart rate variability, and Baevsky stress index are x_{HR} bpm, x_{HRV} ms, and x_{BSI} , respectively. The facial expression is x_{exp} .

Results

1. Subjective Feeling Estimation:

The occupant thinks that the current vehicle's sense of safety is x_{safety} , comfort is x_{comfort} , and travel efficiency is $x_{\text{efficiency}}$.

2. Occupant's Desired Action:

The occupant believes the action the vehicle should take now is x_{action} .

Output

Based on the above information, please analyze the current driving scenario of the autonomous vehicle and the occupant's state, infer the underlying reasons for the occupant's subjective feelings and specific decisions, and generate a clear and logically rigorous chain of thought. The specific content includes:

1. Describe the key visual elements in the driving scenario (such as road type, traffic participants, traffic light status, obstacles, lane lines, weather/lighting conditions, etc.), and determine the vehicle's current behavior.
2. Analyze occupant facial video and physiological states to infer the occupant's emotional state (e.g., tense, relaxed, confused, surprised, etc.).
3. Analyze how the occupant's subjective feelings can be derived from the current vehicle state and occupant expressions and physiological states.
4. Analyze how the occupant's driving decisions can be derived from the current vehicle state and the occupant's subjective feelings.

Requirements

1. Output a chain of thought: Present the complete causal path from environmental perception → emotional/cognitive response → decision-making behavior in the form of a logical chain: "Because... therefore... consequently... ultimately...".
2. Avoid unsubstantiated speculation; all inferences must be based on input information.
3. If certain input is missing or ambiguous, please clearly state this and make cautious inferences based on observable information.
4. Use objective, professional, and concise language, avoiding emotional expressions.
5. Only output the chain of thought; do not include explanatory text.

Here, x_{safety} , x_{comfort} , and $x_{\text{efficiency}}$ are the occupant's subjective feelings of sense of safety, comfort, and travel efficiency, respectively, and x_{action} is the occupant's desired

driving action. Note that the ground truth of the subjective feelings and desired driving decisions are included in the input prompt of the CoT generation process so that the generated CoT samples are correct and valid. Next, the experimenters manually review the generated CoT samples and corresponding video to check if the CoT samples are reasonable and correct. In cases where the generated CoT samples are not reasonable or correct, the experimenters will provide feedback to Qwen3-VL-Plus and ask it to regenerate the CoT samples until valid CoT samples are obtained. In this way, we can ensure that the generated CoT samples are of high quality and correctness, which is crucial for training the VLM in the mental analysis stage. Finally, the CoT samples are screened and refined by the experimenters to make sure they are of high quality and correctness. As for the subjective feeling estimation and decision-making part, we use the participants' annotated subjective feelings and desired driving decisions as the ground truth to train the VLM. The output template of the VLM is formulated as follows:

```
{
    "chain of thought": " $x_{\text{reasoning}}$ ",
    "sense of safety": " $x_{\text{safety}}$ ",
    "comfort": " $x_{\text{comfort}}$ ",
    "efficiency": " $x_{\text{efficiency}}$ ",
    "driving decision": " $x_{\text{action}}$ "
}
```

Note that the JSON format is used to ensure that the output of the VLM is in a structured and machine-readable format, which is easier to process and evaluate.

During training, we use cross-entropy loss as the objective function and AdamW as the optimizer. The learning rate is set to 1×10^{-5} with a linear decay schedule, and the weight decay rate is set to 0.01. The model is trained for 5 epochs with a batch size of 16 on the Bailian Platform of Alibaba Cloud. The training takes about 16 h to complete, and the training cost is approximately 1200 CNY.

3. Results and Discussion

In this section, we evaluate the performance of the proposed ODP paradigm on the developed dataset through comprehensive experiments. Three kinds of experiments are conducted: (1) comparative experiments, where we compare the performance of ODP with the state-of-the-art VLM-based decision-making method; (2) ablation studies, where we validate the effectiveness of the proposed techniques; and (3) case studies, where we use a representative example to illustrate the thinking process of ODP.

3.1. Comparative Experiments

Here, we compare the performance of ODP against the state-of-the-art VLM-based decision-making method, DriveLM [38]. Prior benchmarks have shown that DriveLM can achieve high accuracy in decision-making tasks. Therefore, it is imperative to compare ODP with DriveLM to prove the effectiveness of ODP. Also a Histogram-based Gradient Boosting Classification Tree is included in the comparative experiments to represent rule-based occupant-aware strategies. The rule-based strategy uses not only environmental information but also the ground truth values of subjective feelings in order to facilitate occupant-aware decision-making. Therefore, it should be noted that the use of the ground truth values of subjective feelings will lead to an unfair advantage for rule-based strategy. Moreover, since the participants have annotated their subjective feelings from three aspects (sense of safety, comfort, and travel efficiency) in each scenario, we may assume that their driving decisions are determined by the worst subjective feeling. For example,

if a participant reported a very poor sense of safety, medium comfort, and good travel efficiency, it is very likely that the participant's intention is to enhance safety. As such, we can split all test scenarios into three categories, safety-critical scenarios, comfort-critical scenarios, and efficiency-critical scenarios, according to the relative scores of the subjective feelings. In cases of tied scores, we prioritize safety over comfort, and comfort over efficiency. In total, 1307 safety-critical scenarios, 647 comfort-critical scenarios, and 1096 efficiency-critical scenarios are identified. By doing so, we can further compare the performance of ODP, DriveLM, and rule-based strategy in different types of scenarios to validate the occupant-awareness capability of ODP. Here, we report the decision-making accuracy of ODP, DriveLM, and rule-based strategy on the test set as a whole and in each scenario respectively in Table 1.

Table 1. Comparative results of ODP, DriveLM, and rule-based strategy on the test set.

Model	Test Set Accuracy	Safety-Critical Scenarios Accuracy	Comfort-Critical Scenarios Accuracy	Efficiency-Critical Scenarios Accuracy
Rule-based	29.67%	28.65%	14.11%	34.23%
DriveLM	67.48%	66.64%	67.54%	68.43%
ODP	86.46%	85.77%	86.40%	87.32%

From Table 1, we can see that rule-based methods are incapable of making occupant-aware decisions despite the use of subjective feeling ground truths. Meanwhile, ODP significantly outperforms DriveLM on the test set, achieving an accuracy of 86.46% compared to DriveLM's 67.48%. This demonstrates the superiority of ODP in making occupant-aware driving decisions that can better match the occupant's needs. Furthermore, when we look into different types of scenarios, ODP consistently outperforms DriveLM in safety-critical, comfort-critical, and efficiency-critical scenarios, with accuracies of 85.77%, 86.40%, and 87.32% respectively, compared to DriveLM's 66.64%, 67.54%, and 68.43%. This indicates that ODP is capable of effectively understanding and adapting to the occupant's needs in various driving contexts, further validating its occupant-awareness capability. Additionally, it is interesting to observe that both methods perform best in efficiency-critical scenarios, followed by comfort-critical scenarios, and worst in safety-critical scenarios. This may be because while driving fast is generally considered most efficient, the occupant's perception and understanding of safety is more complex and ambiguous, and thus harder to be fulfilled by AD systems. This observation highlights the needs as well as challenges in developing AD systems that are safer and more trustworthy.

To facilitate a better understanding on the performance of ODP for different categories of meta-actions, we present the confusion matrix of ODP on the test set in Figure 4. The confusion matrix shows that ODP performs well in most categories of meta-actions, achieving 80% or higher accuracies in 12 of the 16 meta-actions. However, it is also noted that ODP is not good at producing certain decisions. The most noteworthy category of meta-action that ODP often overlooks is rapid acceleration, where in 31% of the cases ODP will make an acceleration decision. This phenomenon of refusing to make extreme decisions implies that ODP is conservative in nature. Also there are cases where ODP refuses to go right but goes left instead. This problem is caused by the fact that left turn and left adjustment data are significantly more prevalent than the right turn and right adjustment data. More data should be collected in the future to solve this problem. Overall, the confusion matrix provides a more detailed view of ODP's performance across different meta-action categories, which can help to identify specific areas for improvement in future research.

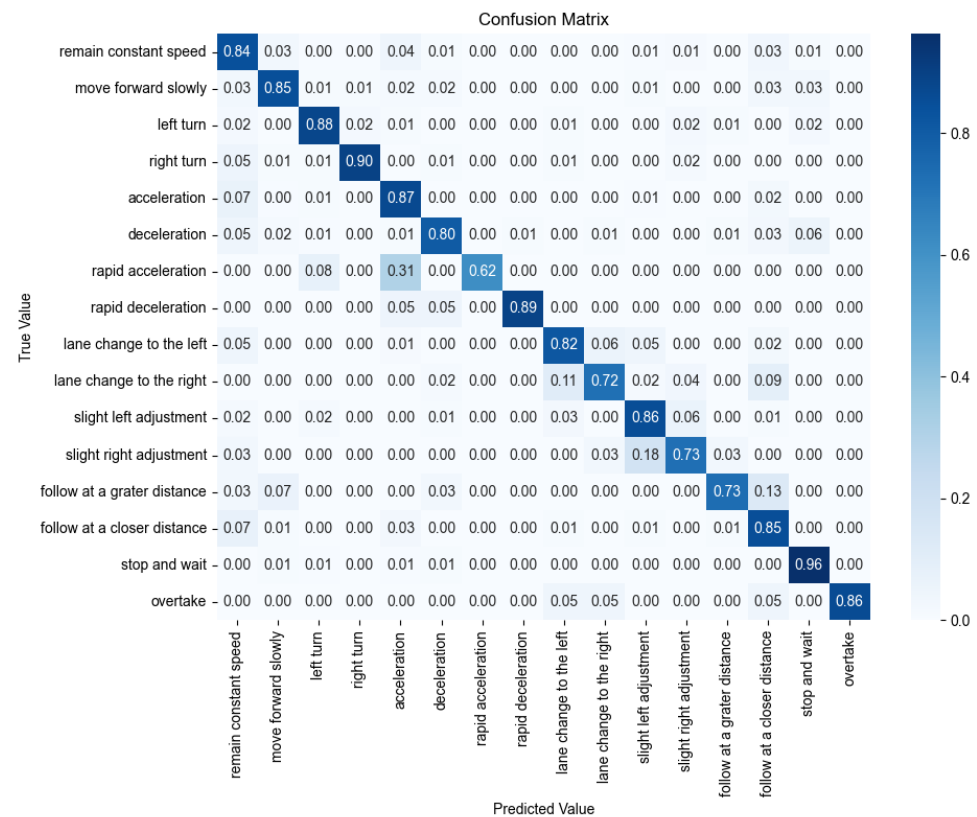


Figure 4. The confusion matrix of ODP on the test set.

3.2. Ablation Studies

To validate the effectiveness of the proposed techniques in ODP, we conduct ablation studies by comparing ODP with its variants. In this study, we build a dataset to fine-tune ODP for the task of occupant-aware decision-making. Moreover, we adopt the CoT reasoning process to enhance ODP's decision-making ability. Therefore, we compare ODP with three variants: (1) ODP without fine-tuning, (2) ODP without occupant information, and (3) ODP without CoT. The experimental results are presented in Table 2.

Table 2. Ablation studies of the proposed techniques on the test set.

Fine-Tuning	Occupant Information	CoT	Test Set Accuracy	Safety-Critical Scenarios Accuracy	Comfort-Critical Scenarios Accuracy	Efficiency-Critical Scenarios Accuracy
			18.95%	17.29%	17.16%	21.99%
✓		✓	51.37%	50.11%	53.01%	51.91%
✓	✓		47.41%	43.69%	50.23%	50.00%
✓	✓	✓	86.46%	85.77%	86.40%	87.32%

From Table 2, we can see that the proposed techniques significantly improve the performance of ODP. Specifically, without fine-tuning, ODP only achieves an accuracy of 18.95% on the test set, which is practically unacceptable. This indicates that the pre-trained VLM alone is insufficient to handle the complex task of occupant-aware decision-making, and that specialized fine-tuning on the developed dataset is essential to adapt ODP for this purpose. Moreover, we also test a variant in which all occupant information is erased from the training data. The environment-only variant yields 51.37% accuracy on the test set, indicating that fine-tuning with the proposed CoT technique, even without occupant information, is helpful to ODP's performance. Additionally, without CoT, ODP achieves an

accuracy of 47.41% on the test set when occupant information is provided. This observation highlights that occupant information is essential for occupant-aware decision-making. However, the accuracies of all these variants are still significantly lower than the full ODP's accuracy of 86.46%. This demonstrates that the CoT technique effectively enhances ODP's reasoning capabilities, enabling it to better analyze the occupant's mental activities and make more accurate driving decisions. Furthermore, similar trends can be observed in different types of scenarios, where both techniques contribute to significant performance improvements. These results validate the effectiveness of the proposed techniques in ODP.

Furthermore, the training configurations of ODP are also validated. In this study, we train Qwen3-VL-8B-Instruct on a mixture of both a manual driving vehicle and an autonomous driving vehicle. The use of a mixture dataset is to ensure that ODP is trained on a diverse range of driving scenarios, including both safe and unsafe situations. This helps to improve the generalization ability of ODP and make it more robust in real-world scenarios. To examine the effectiveness of the data mixture, we build three subsets of the dataset, including (1) Subset-M, which contains 5000 manual driving data; (2) Subset-A, which contains 5000 autonomous driving data; and (3) Subset-X, which contains 2500 manual driving data and 2500 autonomous driving data. We train ODP on each of the subsets, and report the performance in Table 3. The results show that ODP trained on Subset-A, i.e., only autonomous driving data, performed worst among three subsets, indicating that the data collected from autonomous vehicles lack diversity, which is essential for training. ODP trained on Subset-M performs better than Subset-A, highlighting the important role of manual driving data for autonomous driving studies. Finally, ODP trained on Subset-X, which is a mixture of both manual and autonomous driving data, obtains the best performance. This result demonstrates that the mixture dataset can provide more comprehensive and diverse training samples for ODP, which is beneficial for improving the performance of ODP.

Table 3. Performance of ODP trained on different subsets.

Subset	Test Set Accuracy	Safety-Critical Scenarios Accuracy	Comfort-Critical Scenarios Accuracy	Efficiency-Critical Scenarios Accuracy
Subset-M	81.15%	80.03%	82.38%	81.75%
Subset-A	77.66%	75.33%	77.09%	81.22%
Subset-X	83.74%	81.71%	85.01%	85.40%

Moreover, the effect of model architecture and size on performance is also tested. Here, we compare ODP, whose VLM backbone is Qwen3-VL-8B-Instruct, with two other backbones, including Qwen2.5-VL-7B-Instruct and Qwen3-VL-4B-Instruct. The results are presented in Table 4. The results show that Qwen3-VL-8B-Instruct outperforms both Qwen2.5-VL-7B-Instruct and Qwen3-VL-4B-Instruct. Notably, the performance discrepancy between Qwen3 and Qwen2.5 is significantly larger than the discrepancy between larger and smaller models, highlighting that the technological innovations in the field of VLM have greatly enhanced VLMs' capabilities. This demonstrates that the larger and more advanced model architecture helps to enhance the performance of ODP.

Table 4. Performance of ODP with different VLM backbones.

Backbone	Test Set Accuracy	Safety-Critical Scenarios Accuracy	Comfort-Critical Scenarios Accuracy	Efficiency-Critical Scenarios Accuracy
Qwen2.5-VL-7B-Instruct	69.02%	67.52%	69.00%	70.97%
Qwen3-VL-4B-Instruct	81.97%	80.84%	82.63%	83.12%
Qwen3-VL-8B-Instruct	83.74%	81.71%	85.00%	85.40%

3.3. Subjective Feeling Estimation Tests

One key function of ODP is to estimate the subjective feelings of the occupant through mental activity understanding. Here, the subjective feeling estimation performance of ODP is validated on the test set. Here, since the task of subjective feeling estimation is an ordinal regression problem with heavily biased data, we use balanced accuracy, F1 score, and Pearson correlation coefficient r as the evaluation metrics. Unlike accuracy, which is easily inflated by the majority class, balanced accuracy can provide a more truthful evaluation of the model's performance on imbalanced datasets. F1 score is also a useful metric for imbalanced classification problems as it considers both precision and recall, thus providing a more comprehensive evaluation of the model's performance. Moreover, since subjective feelings are ordinal variables, Pearson correlation coefficient r is also calculated to evaluate the correlation between the estimated subjective feelings and the ground truth subjective feelings. The results of all subjective feelings are presented in Table 5, which shows that ODP is able to accurately estimate the subjective feelings of the occupant. Notably, the Pearson correlation coefficient for all subjective feelings exceeds 0.85, highlighting that ODP's estimation of subjective feelings is highly consistent with the ground truth subjective feelings. This demonstrates that ODP can effectively understand the occupant's mental activities and accurately estimate the occupant's subjective feelings, which is essential for making occupant-aware driving decisions.

Table 5. Subjective feeling estimation performance of ODP on the test set.

Subjective Feeling	Balanced Accuracy	F1 Score	r
Sense of safety	0.80	0.81	0.92
Comfort	0.73	0.75	0.89
Travel efficiency	0.79	0.79	0.87

3.4. Case Studies

To illustrate the thinking process of ODP, we take representative examples from the test set as case studies. The driving conditions and the facial image of the occupant are shown in Figure 5. In this case, the ego vehicle is driving at a speed of 50 km/h following the white SUV, which is denoted as the leading vehicle in Figure 5a. Without any warnings, the SUV suddenly slows down and attempts to make a right turn, causing a risky scenario where the ego vehicle may crash into the leading vehicle. In such cases, an AD system may opt to slow down and wait for the leading vehicle. However, ODP is able to observe that the occupant is not concerned about safety, as his facial expression and physiological signals are all steady and relaxed. As such, ODP decides that it is unnecessary to slow down and wait, and instead chooses to perform a small left adjustment to satisfy the participant's need for better efficiency. A snippet of the CoT generated by ODP is shown as follows:

1. Driving Scenario Analysis:

(Omitted for brevity...)

2. Occupant Emotional State Analysis:

- Facial Photo Analysis: The occupant's expression is natural, with a slight smile, and their eyes are looking straight ahead, showing no obvious tension or anxiety. The heart rate is 66.0 bpm, which is within the normal range, further supporting a relaxed emotional state.
- Inference: The occupant's overall emotional state is relaxed and focused, without significant emotional fluctuations.

3. Subjective Perception Derived from Vehicle and Occupant Status:

(Omitted for brevity...)

4. Deriving Driving Decisions from Subjective Perceptions:

- Objective Needs: The vehicle is currently in a complex traffic environment and needs to make minor adjustments to its position based on road conditions to ensure safe passage through obstacles.
- Subjective Perception: Although the occupant experienced some discomfort, he did not show strong emotional fluctuations. Instead, based on his perception of the vehicle's state, he believed the vehicle should perform more precise positional adjustments.
- Decision Derivation: Considering the average travel efficiency, the occupant hoped that the vehicle could adapt to the traffic environment more precisely, ultimately suggesting that the vehicle slightly adjust its position to the left to optimize the driving experience.

After a careful reasoning process, ODP finally reaches the conclusion that the travel efficiency of the vehicle is average, and the occupant's need for precision is not met. Therefore, the vehicle should slightly adjust its position to the left to optimize the driving experience, and the decision for slight left adjustment is made.



Figure 5. A representative example of ODP's decision-making process. (a) Driving conditions captured by the front-facing camera. (b) Facial image of the occupant captured by the facial camera.

In addition, another failure case is shown in Figure 6, which demonstrates how ODP overly relies on the occupant's subjective feelings to make decisions. In this case, the vehicle is making a left turn at a jammed intersection, resulting in the occupant feeling unsatisfactory about travel efficiency. ODP captures the occupant's facial expression and physiological signals, and concludes that the occupant thinks that travel efficiency should be improved. As such, ODP makes the decision of slight left adjustment in order to overtake the key obstacle of the scene, i.e., the red truck. However, since ODP puts its main focus on the occupant, it overlooks the fact that there is a non-drivable area to the left of the vehicle. In fact, the vehicle should make slight right adjustment to obey traffic rules. Consequently, ODP makes the wrong decision in this case. This failure case indicates that there is a trade-off between objective safety and the occupant's subjective feelings, which should be addressed in following studies. A snippet of the CoT generated by ODP is shown as follows:

1. Driving Scenario Analysis:

(Omitted for brevity...)

2. Occupant Emotional State Analysis:

- Facial Photo Analysis: The occupant's facial expression is calm, eyes slightly closed, showing no obvious signs of tension or surprise. The heart rate is 68.0 bpm, within the normal range, and showed no abnormal fluctuations.

- Inference: The occupant is in a relaxed state of mind and does not feel any obvious stress or discomfort.

3. Subjective Perception Derived from Vehicle and Occupant Status:

(Omitted for brevity...)

4. Deriving Driving Decisions from Subjective Perceptions:

- Objective Needs: The vehicle is located below the overpass, in a mixed traffic section, where overall traffic speed is limited.
- Subjective Perception: The occupant is satisfied with the safety and comfort, but perceives potential room for improvement in turning paths and hoped to further enhance travel efficiency.
- Decision Derivation: The occupant believes there is room for improvement in turning path. Video shows other vehicles and obstacles (such as the red truck) on the right side of the vehicle. Ultimately, the occupant thinks the car should now make “slight left adjustment” to improve travel efficiency and optimize the driving experience.

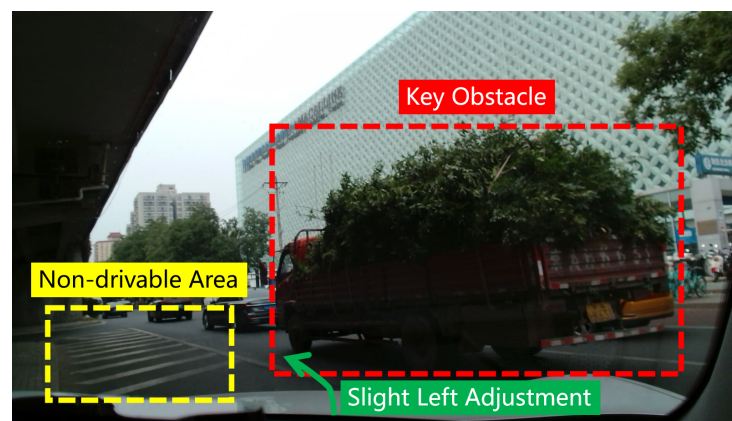


Figure 6. A failure case of ODP.

4. Conclusions

In this study, we propose a novel ODP paradigm that integrates the occupant into the decision-making process of AVs. The proposed paradigm is based on the idea that the occupant's emotional state and subjective perception of the vehicle's status are necessary for making driving decisions that can fulfill the occupant's needs. To this end, we leverage a VLM to analyze the occupant's mental activities through a carefully designed CoT reasoning process, and make driving decisions that can better match the occupant's needs. To support the development and validation of ODP, we curate a large-scale occupant-centric decision-making dataset that includes naturalistic driving data as well as the occupant's states, subjective feelings, and desired driving decisions. Comprehensive experiments on the developed dataset demonstrate that ODP significantly outperforms the state-of-the-art VLM-based decision-making method, validating its effectiveness and occupant-awareness capability. Ablation studies further confirm the effectiveness of the proposed techniques in ODP, including specialized fine-tuning and CoT reasoning. Case studies illustrate the thinking process of ODP in detail, providing insights into how ODP understands and adapts to the occupant's needs. Moreover, since ODP is designed to run entirely using in-vehicle platforms including OMS and AD systems, sensitive biometric data of the occupant will not be leaked to external servers, protecting the privacy of users. Overall, this study highlights the importance of incorporating occupant-awareness into AD systems, paving the way for more personalized, comfortable, and trustworthy autonomous driving experiences.

Since this study is one of the preliminary works that focus on the integration of occupant into the decision-making process of AVs, we acknowledge the following limitations: (1) The proposed paradigm is still in its early stages, and there are many aspects that need to be improved, such as the accuracy of occupant state estimation, the effectiveness of CoT reasoning, and the generalization ability of the model. Also the proposed paradigm may lead to the compromised safety of the AV, which should be rigorously tested before real-world deployment. (2) The developed dataset is limited in scale and diversity, and there are many factors that need to be considered, such as the different driving styles, traffic conditions, and occupant emotions. Moreover, currently, the developed dataset is randomly split, meaning that the same participant will be present in both the training and test sets, which may lead to data leakage. As the scale of the dataset grows, a participant-wise split will be used. (3) Due to the lack of existing data and methods, the experiments conducted in this study are limited, and more comprehensive evaluations are needed to fully validate the effectiveness of the proposed paradigm. (4) Since the VLM is developed on cloud service platforms, it is infeasible to determine if the proposed method is ready for real-world deployment. (5) Cases where the decision-making system should violate the occupant's preferences are not considered in this study, which may lead to safety degradation. Future works may focus on addressing these limitations to further advance the development of occupant-aware AD systems.

Author Contributions: Conceptualization, T.J. and Y.L.; methodology, T.J.; software, T.J.; validation, T.J.; formal analysis, T.J.; investigation, T.J.; resources, X.J. and Y.L.; data curation, X.Z. and T.J.; writing—original draft preparation, T.J.; writing—review and editing, T.J. and Y.L.; visualization, T.J.; supervision, X.J. and Y.L.; project administration, Y.L.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The Independent Research Project of School of Vehicle and Mobility, Tsinghua University.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Tsinghua University Science and Technology Ethics Committee (Medicine) (protocol code: THU01-20250047, date of approval: 2 April 2025).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are not publicly available due to privacy restrictions. The data presented in this study are available on request from the corresponding author.

Acknowledgments: We thank Siteng Jian, Han Dong, He Liu, Yuan Ma, Pengyuan Chu, Jia Zhu, Dehao Kong, Jiangyuan Li, Jiaqi Li, Yingbo Sun, Yuewei Hu, and Zhengyang Han for their help in constructing the dataset. During the preparation of this manuscript, the authors used Doubao-seed-1.6 for the purposes of language editing and proofreading. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

AD	Autonomous Driving
AVs	Autonomous Vehicles
ODP	Occupant-Aware Decision-making Paradigm
OMS	Occupant Monitoring System
VLM	Vision-Language Model
CoT	Chain of Thought
AAA	American Auto Association

LLM	Large Language Model
ADAS	Advanced Driver Assistance System
HVI	Human–Vehicle Interface
IRL	Inverse Reinforcement Learning
ECG	Electrocardiogram
EMG	Electromyograph
EEG	Electroencephalograph
HR	Heart Rate
HRV	Heart Rate Variability
BSI	Baevsky Stress Index
PnC	Planning and Control
NOA	Navigation on Autopilot
SFT	Supervised Fine-tuning

References

1. Sana, F.; Azad, N.L.; Raahemifar, K. Autonomous Vehicle Decision-Making and Control in Complex and Unconventional Scenarios—A Review. *Machines* **2023**, *11*, 676. [\[CrossRef\]](#)
2. Moye, B. *AAA: Fear in Self-Driving Vehicles Persists*; Technical Report; American Auto Association: Menlo Park, CA, USA, 2025.
3. Xiao, J.; Goulias, K.G. Perceived Usefulness and Intentions to Adopt Autonomous Vehicles. *Transp. Res. Part Policy Pract.* **2022**, *161*, 170–185. [\[CrossRef\]](#)
4. Vellenga, K.; Steinhauer, H.J.; Karlsson, A.; Falkman, G.; Rhodin, A.; Koppisetty, A.C. Driver Intention Recognition: State-of-the-art Review. *IEEE Open J. Intell. Transp. Syst.* **2022**, *3*, 602–616. [\[CrossRef\]](#)
5. Karuppasamy, M.; Gangisetty, S.; Rai, S.N.; Masone, C.; Jawahar, C.V. Towards Safer and Understandable Driver Intention Prediction. In Proceedings of the International Conference on Computer Vision, Honolulu, HI, USA, 19–23 October 2025.
6. Trillo, J.R.; Herrera-Viedma, E.; Morente-Molinera, J.A.; Cabrerizo, F.J. A Group Decision-Making Method Based on the Experts' Behavior during the Debate. *IEEE Trans. Syst. Man, Cybern. Syst.* **2023**, *53*, 5796–5808. [\[CrossRef\]](#)
7. Koksalmis, E.; Kabak, Ö. Deriving Decision Makers' Weights in Group Decision Making: An Overview of Objective Methods. *Inf. Fusion* **2019**, *49*, 146–160. [\[CrossRef\]](#)
8. Trillo, J.R.; Herrera-Viedma, E.; Morente-Molinera, J.A.; Cabrerizo, F.J. A Large Scale Group Decision Making System Based on Sentiment Analysis Cluster. *Inf. Fusion* **2023**, *91*, 633–643. [\[CrossRef\]](#)
9. González-Quesada, J.C.; Trillo, J.R.; Porcel, C.; Pérez, I.J.; Cabrerizo, F.J. Modelling Large-Scale Group Decision-Making through Grouping with Large Language Models. *Future Internet* **2025**, *17*, 381. [\[CrossRef\]](#)
10. Carlson, N.R.; Birkett, M.A. *Physiology of Behavior*, 12th ed.; Pearson: London, UK, 2016.
11. Chen, S.; Wang, D.; Zuo, A.; Chen, Z.; Li, W.; Zan, J. Vehicle Ride Comfort Analysis and Optimization Using Design of Experiment. In Proceedings of the 2010 Second International Conference on Intelligent Human-Machine Systems and Cybernetics, Nanjing, China, 26–28 August 2010; Volume 1, pp. 14–18. [\[CrossRef\]](#)
12. Li, J.; Liu, Y.; Ji, X. *Identification of Driver's Braking Intention in Cut-In Scenarios*; SAE Technical Paper 2023-01-0852; SAE International: Warrendale, PA, USA, 2023. [\[CrossRef\]](#)
13. Cheng, Z.; Cheng, Z.Q.; He, J.Y.; Wang, K.; Lin, Y.; Lian, Z.; Peng, X.; Hauptmann, A. Emotion-LLaMA: Multimodal Emotion Recognition and Reasoning with Instruction Tuning. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2024; Volume 37, pp. 110805–110853.
14. Zhu, Y.; Wang, S.; Zhong, W.; Shen, N.; Li, Y.; Wang, S.; Li, Z.; Wu, C.; He, Z.; Li, L. A Survey on Large Language Model-Powered Autonomous Driving. *Engineering* **2025**, *in press*. [\[CrossRef\]](#)
15. Li, J.; Yang, L.; Lv, C.; Chu, Y.; Liu, Y. GLF-STAF: A Global-Local-Facial Spatio-Temporal Attention Fusion Approach for Driver Emotion Recognition. *IEEE Trans. Consum. Electron.* **2025**, *71*, 3486–3497. [\[CrossRef\]](#)
16. Sun, W.; Si, Y.; Guo, M.; Li, S. Driver Distraction Recognition Using Wearable IMU Sensor Data. *Sustainability* **2021**, *13*, 1342. [\[CrossRef\]](#)
17. Dairi, A.; Harrou, F.; Sun, Y. Efficient Driver Drunk Detection by Sensors: A Manifold Learning-Based Anomaly Detector. *IEEE Access* **2022**, *10*, 119001–119012. [\[CrossRef\]](#)
18. Lea, N.A.; Sharmin, S.; Fime, A.A. Drowsiness and Emotion Detection of Drivers for Improved Road Safety. In *Proceedings of the International Conference on Human-Computer Interaction*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 13–26.
19. Ahlström, C.; Kircher, K.; Nyström, M.; Wolfe, B. Eye Tracking in Driver Attention Research—How Gaze Data Interpretations Influence What We Learn. *Front. Neuroergonomics* **2021**, *2*, 778043. [\[CrossRef\]](#) [\[PubMed\]](#)

20. Jiang, T.; Ma, Y.; Zhao, X.; Ji, X.; Liu, Y. NeuralPOS: Physiological Measurement via Remote Photoplethysmography for Driver Monitoring. In Proceedings of the 2024 8th CAA International Conference on Vehicular Control and Intelligence (CVCI), Chongqing, China, 25–27 October 2024; pp. 1–8. [\[CrossRef\]](#)
21. Capallera, M.; Angelini, L.; Meteier, Q.; Khaled, O.A.; Mugellini, E. Human-Vehicle Interaction to Support Driver’s Situation Awareness in Automated Vehicles: A Systematic Review. *IEEE Trans. Intell. Veh.* **2023**, *8*, 2551–2567. [\[CrossRef\]](#)
22. Zhou, S.; Lan, R.; Sun, X.; Bai, J.; Zhang, Y.; Jiang, X. Emotional Design for In-Vehicle Infotainment Systems: An Exploratory Co-Design Study. In Proceedings of the HCI in Mobility, Transport, and Automotive Systems, Virtual Event, 26 June–1 July 2022; pp. 326–336. [\[CrossRef\]](#)
23. Li, J.; Liu, Y.; Ji, X.; Tao, S. *Detection of Driver’s Cognitive States Based on Lightgbm with Multi-Source Fused Data*; SAE Technical Paper 2022-01-0066; SAE International: Warrendale, PA, USA, 2022. [\[CrossRef\]](#)
24. Lin, C.; Zhu, X.; Wang, R.; Zhou, W.; Li, N.; Xie, Y. Early Driver Fatigue Detection System: A Cost-Effective and Wearable Approach Utilizing Embedded Machine Learning. *Vehicles* **2025**, *7*, 3. [\[CrossRef\]](#)
25. Lin, X.; Huang, Z.; Ma, W.; Tang, W. EEG-based Driver Drowsiness Detection Based on Simulated Driving Environment. *Neurocomputing* **2025**, *616*, 128961. [\[CrossRef\]](#)
26. Xu, X.; Yao, B.; Dong, Y.; Gabriel, S.; Yu, H.; Hendler, J.; Ghassemi, M.; Dey, A.K.; Wang, D. Mental-LLM: Leveraging Large Language Models for Mental Health Prediction via Online Text Data. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*; Association for Computing Machinery: New York, NY, USA, 2024; Volume 8, pp. 1–32. [\[CrossRef\]](#)
27. Hu, J.; Dong, T.; Luo, G.; Ma, H.; Zou, P.; Sun, X.; Guo, D.; Yang, X.; Wang, M. PsychoLLM: Enhancing LLM for Psychological Understanding and Evaluation. *IEEE Trans. Comput. Soc. Syst.* **2025**, *12*, 539–551. [\[CrossRef\]](#)
28. Hasenjaeger, M.; Heckmann, M.; Wersing, H. A Survey of Personalization for Advanced Driver Assistance Systems. *IEEE Trans. Intell. Veh.* **2020**, *5*, 335–344. [\[CrossRef\]](#)
29. Marina Martinez, C.; Heucke, M.; Wang, F.Y.; Gao, B.; Cao, D. Driving Style Recognition for Intelligent Vehicle Control and Advanced Driver Assistance: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 666–676. [\[CrossRef\]](#)
30. Wu, J.; Yan, Y.; Liu, Y.; Liu, Y. Research on Anthropomorphic Obstacle Avoidance Trajectory Planning for Adaptive Driving Scenarios Based on Inverse Reinforcement Learning Theory. *Engineering* **2024**, *33*, 133–145. [\[CrossRef\]](#)
31. Yang, H.; Zhou, Y.; Wu, J.; Liu, H.; Yang, L.; Lv, C. Human-Guided Continual Learning for Personalized Decision-Making of Autonomous Driving. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 5435–5447. [\[CrossRef\]](#)
32. Wang, X.; Guo, Y.; Ban, J.; Xu, Q.; Bai, C.; Liu, S. Driver Emotion Recognition of Multiple-ECG Feature Fusion Based on BP Network and D-S Evidence. *IET Intell. Transp. Syst.* **2020**, *14*, 815–824. [\[CrossRef\]](#)
33. Jiang, T.; Li, J.; Ma, L.; Ji, X.; Liu, Y. Passenger Comfort Assessment via Motion Complexity Analysis for Autonomous Vehicles. *Chin. J. Mech. Eng.* **2025**, *38*, 149. [\[CrossRef\]](#)
34. Thirunavukkarasu, G.S.; Abdi, H.; Mohajer, N. A Smart HMI for Driving Safety Using Emotion Prediction of EEG Signals. In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 October 2016; pp. 004148–004153. [\[CrossRef\]](#)
35. Liu, H.; Li, C.; Wu, Q.; Lee, Y.J. Visual Instruction Tuning. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 10–16 December 2023.
36. OpenAI. GPT-4V(Ision) System Card. Available online: https://cdn.openai.com/papers/GPTV_System_Card.pdf (accessed on 10 January 2026).
37. Bai, J.; Bai, S.; Yang, S.; Wang, S.; Tan, S.; Wang, P.; Lin, J.; Zhou, C.; Zhou, J. Qwen-VL: A Versatile Vision-Language Model for Understanding, Localization, Text Reading, and Beyond. *arXiv* **2023**, arXiv:2308.12966.
38. Sima, C.; Renz, K.; Chitta, K.; Chen, L.; Zhang, H.; Xie, C.; Luo, P.; Geiger, A.; Li, H. DriveLM: Driving with Graph Visual Question Answering. In Proceedings of the European Conference on Computer Vision, Paris, France, 2–3 October 2023.
39. Tian, X.; Gu, J.; Li, B.; Liu, Y.; Hu, C.; Wang, Y.; Zhan, K.; Jia, P.; Lang, X.; Zhao, H. DriveVLM: The Convergence of Autonomous Driving and Large Vision-Language Models. In Proceedings of the Conference on Robot Learning, Munich, Germany, 6–9 November 2024.
40. Xie, S.; Kong, L.; Dong, Y.; Sima, C.; Zhang, W.; Chen, Q.A.; Liu, Z.; Pan, L. Are VLMs Ready for Autonomous Driving? An Empirical Study from the Reliability, Data and Metric Perspectives. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Honolulu, HI, USA, 19–23 October 2025; pp. 6585–6597.
41. Baevsky, R.M.; Chernikova, A.G. Heart Rate Variability Analysis: Physiological Foundations and Main Methods. *Cardiometry* **2017**, *10*, 66–76. [\[CrossRef\]](#)
42. Bai, S.; Cai, Y.; Chen, R.; Chen, K.; Chen, X.; Cheng, Z.; Deng, L.; Ding, W.; Gao, C.; Ge, C.; et al. Qwen3-VL Technical Report. *arXiv* **2025**, arXiv:2511.21631. [\[CrossRef\]](#)

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.