Undergraduate Project Report
2023/24

Design and Development of a Human-Agent Collaboration Model for Situation Awareness in Cockpit

Name:	Yitong Hu
School:	International School
Class:	2020215111
QMUL Student No.:	200980434
BUPT Student No.:	2020213350
Programme:	e-Commerce Engineering with Law

Date: 25-4-2024

Table of Contents

Abstract

The advent of autonomous driving technology brings significant advancements and challenges, particularly in achieving seamless human-agent collaboration (HAC) at various levels of automation. This paper introduces HarmonyCockpit (HCockpit), a sophisticated agent architecture designed to enhance situational awareness (SA) in automotive environments. HCockpit integrates cutting-edge large multimodal models (LMMs) to facilitate effective communication and collaboration between human drivers and autonomous driving systems. By synthesizing insights from both the vehicle's internal and external environments, HCockpit generates context-aware tasks in natural language, which are then translated into actionable directives within a predefined cockpit model. This orchestration aids in aligning human and agent actions towards coordinated outcomes, thereby improving the driving experience and safety. Experimental validations carried out in a simulated environment (GTAV) demonstrate HCockpit's ability to reduce cognitive load and augment situational awareness for drivers, particularly in complex driving scenarios. The architecture not only showcases the potential of LMMs in practical applications but also advances the field of human-agent collaboration in autonomous driving.

Keywords

human-agent collaboration, context awareness, situation awareness, implicit interaction

摘要

随着自动驾驶技术的发展，如何在不同自动化水平上实现人机无缝协作（HAC）成为一大挑战和进步。本文介绍了 HarmonyCockpit（HCockpit），这是一个先进的代理架构，旨在提高汽车环境中的情境感知（SA）。HCockpit 整合了尖端的大型多模态模型（LMMs），以促进人类驾驶员与自动驾驶系统之间的有效沟通和协作。通过综合车辆内部和外部环境的洞察，HCockpit 生成了自然语言级别的上下文感知任务，随后将这些任务转化为预定义驾驶舱模型中的可执行指令。此种协调有助于将人类和代理的行动对齐，从而改善驾驶体验和安全性。在模拟环境（GTAV）中进行的实验验证表明，HCockpit 能够减轻驾驶员的认知负担并增强其在复杂驾驶场景中的情境感知能力。该架构不仅展示了 LMMs 在实际应用中的潜力，还推动了自动驾驶中人机协作领域的发展。

关键词

人机协作、上下文感知、态势感知、隐式交互

Chapter 1: Introduction

As the era of fully autonomous driving approaches, autonomous driving systems still require time to handle diverse situational challenges, environmental variables, and unforeseen circumstances. Prior to full automation, this process will be gradual, signifying the necessity for human supervision. Meanwhile, as autonomous driving technology evolves, it introduces a range of new challenges. Currently, many original equipment manufacturers (OEMs) are adopting Level 2+ or Level 3 autonomous driving features, which allow drivers to temporarily relinquish control of specific driving functions, thus balancing vehicle performance and cost-effectiveness.

However, these systems may intermittently require human re-engagement in vehicle operation; conversely, drivers may need the support of automatic systems in certain scenarios. At this level of autonomous driving, achieving smooth communication and collaboration between drivers and automated systems is crucial to enhancing the driving experience and safety. First, establishing an efficient, transparent communication mechanism between automated systems and human drivers to ensure rapid and accurate information exchange at critical moments is a significant challenge. Second, how to process and integrate information from various sensors and data sources to achieve accurate situational awareness is a key issue in system design.

This project explores a sophisticated interaction system, HarmonyCockpit (HCockpit), aimed at promoting situational awareness by combining insights from the vehicle's internal and external environments, serving as a bidirectional link between the automated system and the driver, and coordinating actions to facilitate context-aware human-machine collaboration tasks. HCockpit is a novel agent architecture designed to incorporate cutting-edge large multimodal models (LMMs) and orchestrate human-agent collaboration (HAC) in the cockpit with transparency. HCockpit achieves situational awareness by integrating insights into both the internal and external environments of the cockpit and generates natural language-level tasks in context-awareness to facilitate HAC. Subsequently, HCockpit translates tasks into actions adapting to a predefined cockpit model, thereby guiding both humans and agents towards coordinated action. Metaphorically speaking, within a cockpit setting that includes a human driver and an autonomous driving system (agent), HCockpit serves as the orchestrating force between human and agent.

In designing HCockpit, this paper conducted a series of experiments in the GTAV simulation environment to test and validate the system's effectiveness. Experimental results indicate that HCockpit effectively enhances the situational awareness of drivers and reduces their cognitive load, especially in complex or emergency driving situations. The HCockpit architecture not only demonstrates the feasibility of achieving effective human-machine collaboration in advanced autonomous driving environments but also proves the potential of large multimodal models in practical applications. Through this project, the paper not only enhances the driving experience and safety but also provides robust technical support for addressing attention concentration issues.

The following sections of this paper will detail the research background, design, and implementation of HCockpit, including the system architecture, key technological components, and the experimental setup and results in the GTAV simulation environment. Finally, the paper will discuss the practical applications of HCockpit and future research directions.

Chapter 2: Background

2.1 Situation Awareness

2.1.1 Introduction

Situation awareness (SA) is the perception of elements in the environment, comprehension of their meaning, and the projection of their status in the near future. It involves being aware of the start and end of situations, as well as any active situation at any given time. SA in intelligent cockpit refers to the ability of an autonomous vehicle to perceive its environment, understand the significance of the perceived information, and predict future states of the environment.

SA is a critical aspect of autonomous driving, as it directly impacts the safety and efficiency of the vehicle's operation. One main issue in SA includes attentional tunneling, where operators focus on a single goal and lose awareness of the overall picture, and stimuli that may divert attention from important aspects, leading to erroneous decisions (1).

2.1.2 Representative Works

The development of research hotspots in this field has been focused on improving the SA of autonomous vehicles through various methods, including multimodal sensing, machine learning, and interface design.

Multimodal Sensing for SA (2) demonstrated a multimodal sensing approach for objective SA monitoring in autonomous driving. The study used physiological sensor data from electroencephalogram and eye-tracking to assess SA. The results showed that a multi-physiological sensor-model outperformed the single sensing model, suggesting that multimodal sensing can objectively predict SA.

Machine Learning for SA Research by (3) used supervised machine learning techniques for both reactive predictions (short-term) and motivation-based predictions (long-term) to predict the behavior of other traffic participants and decide what to do with these. The study showed that even simple prediction and decision algorithms can considerably improve the current status quo, although more advanced models increase complexity.

Interface Design for SA (4) explored the design of the vehicle terminal interface in a closed dark cabin driving environment to improve the driver's perception of the environmental information outside the cabin and the ease of use of the interface. The study found that the design method effectively enhanced the driver's SA.

Attention-guiding Techniques for SA (5) proposed to improve drivers' takeover performance by utilizing attention-guiding techniques when delivering the takeover request (TOR) in semi-autonomous driving. The preliminary experiment indicated that this method reduced drivers' collision rate and mental workload.

2.1.3 Challenges

Despite the advancements in SA for autonomous driving, several challenges remain:

1. Interface Design: The design quality of the vehicle terminal interface directly affects the driver's SA level during driving. Therefore, improving the driver's perception of the environmental information and the ease of use of the interface is a challenge (4).

2. Human-Automation Interaction: A major question in human-automation interaction is whether tasks should be traded or shared between human and automation. This dilemma may impact the design of automation systems (6).

3. Traffic Sign Detection: Traffic sign detection and recognition is a critical aspect of the environmental awareness module of autonomous driving. Early traffic sign recognition methods were mostly based on color features, shape features, or multi-feature fusion (7).

4. Transparency of agents: In the context of automated vehicles, the transparency and reliability of in-vehicle intelligent agents significantly impact driver perception, workload, and SA (8).

2.2 Human-Agent Collaboration

2.2.1 Introduction

Human-Agent Collaboration (HAC) in intelligent cockpit refers to the interaction between human drivers and autonomous driving systems. This field aims to enhance the safety, efficiency, and user experience of autonomous vehicles by leveraging the strengths of both human drivers and autonomous systems. The main issues in this field include path planning, perception of the dynamic world, decision-making, and communication between human drivers and autonomous systems (Agapito and Fallon, 2022; Khemchandani et al., 2023; Plebe et al., 2022).

2.2.2 Representative Works

Applications of Large-Scale Foundation Models for Autonomous Driving (Huang et al., 2024) investigates the application of large language models (LLMs) and foundation models in autonomous driving. The authors propose that these models can be used to reformulate autonomous driving by leveraging human knowledge, common sense, and reasoning. The models can be applied in various areas, including simulation, world model, data annotation, and planning or end-to-end solutions.

(Liao et al., 2023) introduces a sophisticated encoder-decoder framework, the Context-Aware Visual Grounding (CAVG) model, to address visual grounding in autonomous vehicles. The model integrates five core encoders with a Multimodal decoder, enabling it to capture contextual semantics and learn human emotional features. The model demonstrated high prediction accuracy and operational efficiency, even with limited training data.

In conclusion, multi-modal large-scale models, such as GPT-4, have been applied in autonomous driving to enhance human-agent collaboration. These models can process and interpret a range of cross-modal inputs, yielding a comprehensive understanding of the correlation between verbal commands and corresponding visual scenes. They can also learn human emotional features, which can be useful in understanding and responding to human drivers' intentions and emotions (Cui et al., 2023; Liao et al., 2023; L. Wang et al., 2023; Z. Yang et al., 2023).

Real Time Human Assisted Path Planning for Autonomous Agent using VR (Khemchandani et al., 2023) focuses on path planning, a critical aspect of autonomous driving. The researchers developed a virtual reality (VR) system to train defense personnel in path planning for various operations in remote areas. The system simulates real-world scenarios, including traffic light systems, AI car navigation algorithms, and rescue operations, providing a cost-effective and safe training environment.

Toward Policy Explanations for Multi-Agent Reinforcement Learning:

(Boggess et al., 2022) presents novel methods to generate policy explanations for multi-agent reinforcement learning (MARL), a technique used in autonomous driving. The authors developed methods to summarize agent cooperation and task sequence and to answer queries about agent behavior. The study found that these explanations improved user performance and satisfaction.

Distributed cognition for collaboration between human drivers and self-driving cars: This paper proposes a collaboration mechanism based on the concept of distributed cognition. The authors suggest that intelligence lies not only in the individual entities (human or autonomous agent) but also in their interaction. The study uses a driving simulator to demonstrate the collaboration in action, showing how the human can communicate and interact with the agent in various ways with safe outcomes (Plebe et al., 2022).

2.2.3 Challenges

Despite the progress made in human-agent collaboration in autonomous driving, several challenges remain. These include accurately representing the mutual effects of vehicles and modeling dynamic traffic environments in mixed autonomy traffic, which includes both autonomous vehicles and human-driven vehicles (Liu et al., 2022). Another challenge is managing the risk that an agent's action could harm a friendly computer, which must be balanced against the losses that could occur if the agent does not act (Kott, 2023). Lastly, there is a need for more research on how to maintain human expertise and relevance in professional decision-making as automation increases (X. Li et al., 2023).

2.3 Summary

Addressing the challenges highlighted above, this project introduces HarmonyCockpit (HCockpit), a framework that integrates advanced multi-modal large-scale models to facilitate transparent human-agent collaboration (HAC) within the cockpit environment. HCockpit cultivates situational awareness by synthesizing information from both the cockpit's internal and external milieus and directs actions in concert with established cockpit functions to support HAC tasks grounded in situational cognizance.

To assess HCockpit's efficacy and derive insights, the HarmonyCopilot (HCopilot) was developed as an operational example of the HCockpit framework, utilizing cutting-edge multi-modal large-scale models alongside conventional intelligent cockpit designs. As a reciprocal link between the autonomous driving system and the driver, HCopilot strives to augment the driving experience and safety via an integrated human-vehicle interface.

Different from (Huang et al., 2023; Liao et al., 2023; L. Wang et al., 2023; S. Wang et al., 2023; Wen et al., 2023)s' works, this project research introduces the HCockpit framework and the HCopilot exemplar as pioneering contributions to autonomous driving technology, emphasizing AI-driven collaboration between humans and machines. It notably accentuates situational awareness and undertakes passive human-system interaction. Leveraging substantial multi-modal models, the initiative endeavors to enrich comprehension of both vehicular confines and the external environment, thereby enhancing the response capability and transparency of the autonomous system. Departing from conventional autonomous driving studies, it incorporates advanced functionalities like semantic comprehension, driver intent prediction, and bidirectional communication—innovations that position it at the forefront of the field.

HCockpit notably excels in personalizing user experience by proactively adapting to the driver's behaviors and preferences, enhancing trust and satisfaction with the system. In terms of safety, the model responds promptly to lapses in the driver's focus or when faced with challenging driving scenarios that surpass the autonomous system's capacity, proactively signaling the driver to assume control. This feature serves to avert potential accidents and bolsters overall driving safety. Such advancements highlight HCockpit's role in not only improving autonomous driving performance but also in offering a tailored driving experience with significant market potential.

Chapter 3: Design and Implementation

This section progressively introduces the background, design, and implementation of HCockpit. Initially, the Preliminaries section introduces the basic concepts and terminologies involved in this paper and provides foundational assumptions and modeling. Subsequently, the HCockpit Architecture section systematically presents the modular architecture of HCockpit and explains the design rationale behind it. Finally, the HCopilot Implementation section details the specific implementation of HCopilot (an AI copilot developed following HCockpit architecture).

3.1 Preliminaries

3.1.1 Environment, Entity, and Agent

The environment consists of all entities within it, where entities interact with each other. However, entities have varying levels of intelligence; some can only passively respond to stimuli from the environment, while others can actively observe the environment and form a certain understanding, then autonomously plan and execute a series of actions to pursue specific goals. To distinguish between these, this paper refers to the second type of entity as an agent. From the agent's perspective, all other entities in its environment are perceivable or controllable objects. Naturally, the agent's ability to perceive or control other objects depends on its capabilities.

This paper defines the global environment as consisting of the external environment outside the ego vehicle and the internal environment. The external environment includes entities like the road, traffic, and other vehicles, while the internal environment consists of devices defined by the cockpit model and three types of agents: 1. autonomous driving system (AI driver), 2. human driver, and 3. AI copilot (an agent developed following HCockpit architecture). The first two are considered independent homomorphic agents, meaning they have the same observation space (the external environment). Unlike them, the AI copilot has global observation, meaning it can perceive all entities in both the internal and external environments.

3.1.2 Action

Action refers to the control executed by an agent on entities in the environment, representing the final step in achieving its goals or responding to environmental changes. In an environment with multiple agents, from the ego agent's perspective, unlike directly controlling an entity, it may be challenging to directly control other agents; instead, it needs to communicate its intentions to others to achieve a certain goal.

In this paper, since the AI driver and human driver are homomorphic agents, they share the same action space, which is the control of the vehicle's motion controller. The AI copilot (considered as the ego agent in this paper), however, has a different action space, which is to control the devices, including the motion controller, in the cockpit.

It is noteworthy that in real driving scenarios, human drivers might control not only the motion controller but also improve the driving experience by controlling cockpit devices such as the dashboard, display screen, and air conditioning. The HCockpit architecture initially considered such scenarios, but to simplify the problem for rapid validation of HCockpit's usability, this paper restricts the human driver's actions to controlling the motion controller. Moreover, this paper does not consider the scenario where the human driver and AI driver jointly control the motion controller, as this often exacerbates the human driver's driving burden, which contradicts the initial intention of this paper.

3.1.3 Human-Agent Collaboration

Human-Agent Collaboration (HAC) refers to the process where both parties achieve a specific goal in a shared environment. In this paper, HAC is the ultimate goal of the AI copilot, referring to the AI driver and human driver jointly achieving a safe, comfortable, and efficient driving experience in a shared driving environment. The HAC goals focused on in this paper include: 1. Utilizing the AI driver's environmental perception capabilities to enhance the driver's situational awareness, 2. Combining the human driver's state to provide driving style suggestions to the AI driver. To achieve the HAC goals, the AI copilot needs to decompose this complex goal into a series of simple sub-tasks (HAC tasks), such as: 1. Controlling relevant hardware in the cockpit to send voice reminders or video streams to the driver, 2. Providing control suggestions or takeover requests to the AI driver based on the driver's situational awareness. Since the AI copilot does not have a physical entity, cockpit devices will serve as the tools for executing HAC tasks (embodied), requiring the AI copilot to master the control of cockpit devices.

3.1.4 Cockpit Model

This paper introduces the cockpit model as a simplified model of the cockpit environment. The cockpit model defines the functions of all devices within the cockpit and the control permissions of different roles (agents). Technically, the cockpit model provides a unified application programming interface (API). This allows the AI copilot to finely use the functions of cockpit devices to complete HAC tasks. It is worth noting that the cockpit model does not involve the specific control details of devices but has designed a simple device control interface for evaluation and demonstration.

3.1.5 Large Multimodal Model

Large Multimodal Model (LMM) refers to the general term for generative pre-trained deep learning models that can handle multiple input modalities (such as text, images, etc.) and generate text outputs. The input to an LMM is usually referred to as ”prompt,” which contains the task description and input data for the model.

In this paper, LMMs are used to implement the core functions of the AI copilot, including situational awareness, task orchestration, and control. Thanks to LMM's broader general knowledge and advanced reasoning capabilities, the AI copilot can understand complex traffic environments and the intentions of human drivers in real driving scenarios, thereby appropriately assisting human drivers. As of April 2024, the most advanced LMM is developed by OpenAI, GPT-4 Turbo, which can be accessed via the OpenAI API.

3.2 HCockpit Architecture

Figure 1 shows the overall architecture of HCockpit, which models three types of agents (human driver, AI driver, and AI copilot) and two types of entities (cockpit devices and external entities). Specifically, the human driver and AI driver perceive information from cockpit devices and external entities, then control the motion controller within cockpit devices. It is noteworthy that their perceptions and control methods of entities differ. The human driver perceives cockpit devices and external entities through visual, auditory, and tactile means, then directly physically controls the motion controller; the AI driver perceives external entities using onboard sensors and interacts with cockpit devices (perception and control) according to the API defined by the cockpit model.

For the AI copilot, it perceives the behaviors of the human driver and AI driver through cockpit devices and collects observations from the AI driver on external entities, aggregating them to obtain a global observation of the environment. Based on this global observation, the AI copilot's goal is to control cockpit devices to communicate with the human driver and AI driver to complete HAC tasks. To achieve this goal, the text hierarchically designs four modules for the AI copilot in HCockpit: the situation awareness module, planning module, memory module, and control module, which will be introduced one by one below.

3.2.1 Situation Awareness Module

The AI copilot needs to continuously observe the global environment (external + internal) to achieve situational awareness (SA). For the internal environment, this paper only considers the cockpit's main driver. Devices like cameras, eye trackers, microphones, and motion controllers in the cockpit will continuously record the behavior of the human driver. Through communica-

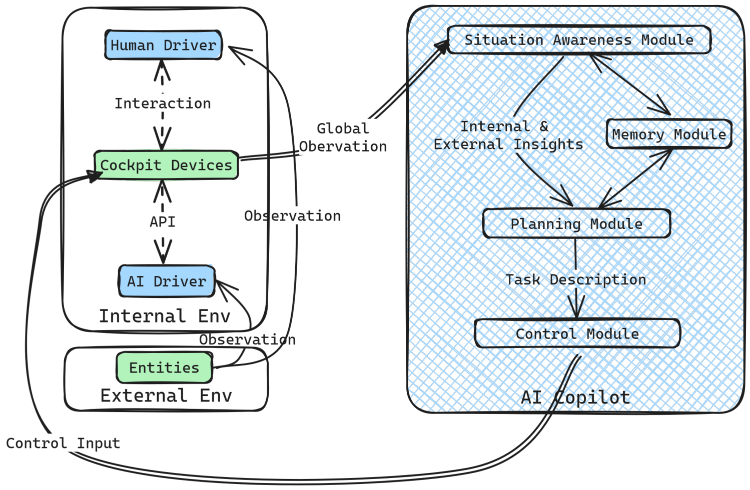

Figure 1: HCockpit Architecture

tion with cockpit devices, the SA module can receive multimodal data inputs containing natural language instructions, physiological and control data from the driver, use LMM to extract semantic information and driver intentions from the scene, and then save it as structured state text as an observation of the internal environment. It should be noted that the SA module does not observe the state of the AI driver because when the human driver predominantly controls the vehicle, there is no need to consider the AI driver's strategy or intentions; conversely, when the AI driver predominantly controls, the human driver's intentions or preferences are crucial.

For the external environment, the SA module obtains observations of external entities through the AI driver's sensors and uses LMM to obtain interpretable environmental semantic information and projections of future states and events. Similarly, the SA module saves the above information as state text, serving as an observation of the external environment.

Subsequently, the SA module aggregates the two observations into structured global observation state text, allowing the AI copilot to further analyze and make decisions.

3.2.2 Planning Module

The planning module, combining comprehensive SA and context of the global environment, orchestrates tasks to facilitate HAC. Specifically, the planning module receives input of global observation state text, then analyzes the human driver's goals or needed assistance, identifies specific HAC goals, and arranges HAC tasks. During this process, the planning module follows the principle of prioritizing driving safety followed by driving experience.

The opacity of intelligent system decisions can hinder the establishment of human-agent trust, potentially leading human drivers to question the system's decisions, leading to attentional tunneling and affecting driving safety. The design of the planning module takes this into account; driven by LMM, its HAC goals and tasks are described in natural language, which is easy to understand and helps enhance human-agent trust and driving safety.

3.2.3 Memory Module

Since the AI copilot's assistance is a continuous process, context awareness is crucial. Inspired by ChatGPT, this project designed the memory module for HCockpit, which separately saves the historical inputs and outputs of the SA module and planning module and uses custom summarization techniques to provide appropriate context information for LMM, enabling the AI copilot to provide more comprehensive task orchestration.

3.2.4 Control Module

The control module is the execution module of the AI copilot, which receives natural language level HAC tasks generated by the planning module and converts them into control codes for cockpit devices through the API defined by the cockpit model. The control module is also driven by LMM, essentially a text-to-code converter that combines external knowledge bases (API document).

From the perspective of automation and control systems, the control module can be seen as a controller, where its input is HAC tasks (reference), the output is control codes for cockpit devices (control input), and the various devices in the cockpit serve as the plant, forming a simple feedforward control system.

3.3 HCopilot Implementation

To assess the efficacy of HCockpit and garner deeper understandings, this paper developed HarmonyCopilot (HCopilot) as an AI copilot agent following the HCockpit architecture and tested HCopilot in the Grand Theft Auto V (GTAV) simulation environment.

HCopilot employs the state-of-the-art LMM (GPT-4 Turbo) and integrates a demonstrative intelligent cockpit model and tools API for controlling vehicle cockpit devices in GTAV. Internally, HCopilot adopts the design of LMM as an agent, meaning each module (except the memory module) is an independent LMM instance. Therefore, this paper uses the LangChain framework to implement the information transfer between modules and module-to-module communication, as well as automatic context summarization.

3.3.1 Tools API

GTAV is a globally renowned open-world game that realistically recreates the urban environment and traffic scenes of Los Angeles, providing players with a realistic first-person driving

Figure 2: First-Person Driving in GTAV

experience, as shown in Figure 2. ScriptHookVDotNet (SHVDN) is a runtime that interfaces between custom .NET code and GTAV, running under Script Hook V to utilize GTAV script native functions in custom *.asi plugins.

This paper defines a basic functionality cockpit model for GTAV and develops cockpit device APIs based on SHVDN and other packages. From the perspective of HCopilot, these APIs are referred to as tools API, which can be invoked by HCopilot. Specifically, the tools API includes control and perception functions:

1. `set_cam_to()`: Controls the external camera of the ego vehicle to lock onto a specific entity or location and displays the video stream on the Head-Up Display (HUD).
2. `set_speech()`: Controls the cockpit's speakers to read aloud a specific text.
3. `obs_eye()`: Obtains the video stream of the player (human driver) through a network camera, then uses the Beam Eye Tracker and OpenTrack to obtain the human driver's eye movement data, including gaze coordinates, returning array format data.
4. `obs_hctrl()`: Communicates with GTAV via SHVDN to obtain the human driver's control data, including keyboard input, returning JSON format data.
5. `obs_sensor()`: Uses SHVDN to obtain external environmental data of the ego vehicle, including information on vehicles within a 50-meter radius (relative coordinates, speed, acceleration, and collision box, including the ego vehicle), returning JSON format data.
6. `obs_hview()`: Uses SHVDN to obtain the first-person view image of the human driver, returning png format image.

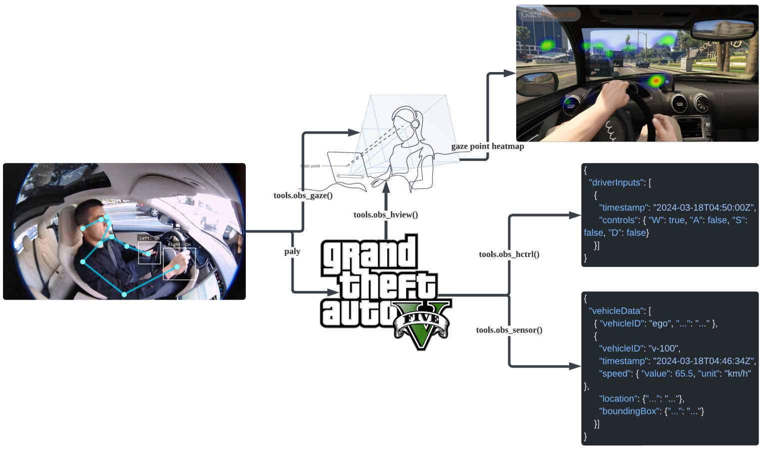

Figure 3: Data Collection for HCopilot

3.3.2 Data Collection

HCopilot calls the `obs_eye()` function at a frequency of 30 Hz to obtain the gaze coordinates of the human driver and generates heatmaps at a frequency of 0.5 Hz. Simultaneously, HCopilot calls the `obs_hctrl()`, `obs_sensor()`, and `obs_hview()` functions at a frequency of 0.5 Hz and overlays the gaze point heatmap on the first-person view image. Figure 3 fully demonstrates the data collection process of HCopilot.

3.3.3 Prompt Engineering

Prompt, also known as the input to the LMM, is the description of the model input data and the task. For OpenAI's GPT-4 Turbo model, the commonly used prompt consists of a system prompt and a user prompt. The system prompt is the initial input to the model, used to guide or set the background of the dialogue. The user prompt is the information entered by the user, used to guide the model to provide specific answers or perform tasks. The design of the prompt directly affects the quality of the LMM's output, so it needs to be carefully designed and continuously optimized, a process known as prompt engineering. OpenAI's prompt engineering guide proposes the following tactics:

1. Include detailed information in the query to obtain more relevant answers.
2. Ask the model to adopt a persona.
3. Use delimiters to clearly indicate different parts of the input.
4. Follow the chain-of-thought (CoT) prompting, specifying the steps required to complete the task.
5. Provide examples.

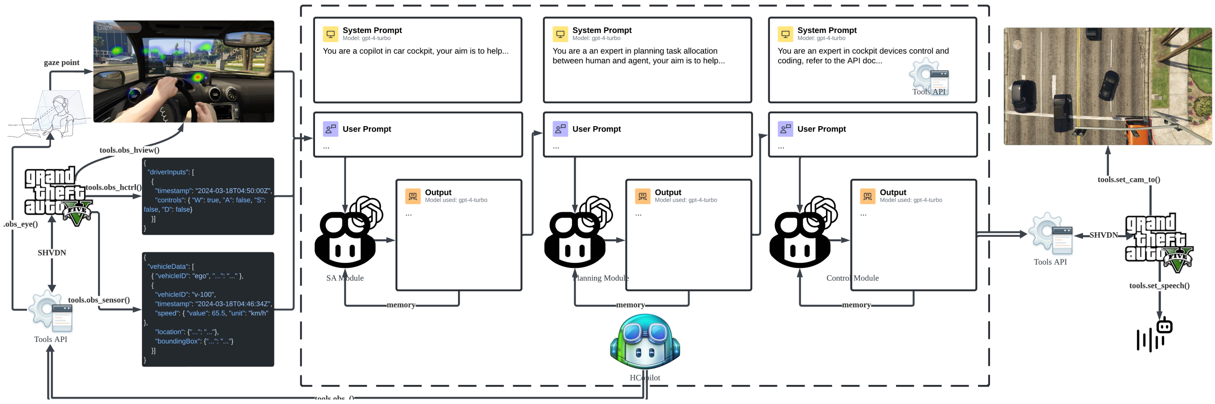

Figure 4: Overall Workflow of HCopilot

The SA module, planning module, and control module prompts of HCopilot all follow the above tactics to ensure the quality of LMM's output. For example, in the SA module, for tactic.1, the system prompt mentions: ”You are the AI copilot in a car cockpit environment...”. For tactic.4, the system prompt specifies the steps required to complete the task: ”1. Analyse the human driver's eye gaze and control data, 2. Observe the external environment, 3. Conduct a comprehensive analysis...”. The complete prompts for the three modules are shown in Additional Appendices.

3.3.4 Overall Workflow

Figure 4 shows the overall workflow of HCopilot. In the SA module, GPT-4 Turbo processes the input multimodal data, extracting semantic information and driver intentions, and saves these as structured state text. The memory module then stores this state text, providing appropriate context for the LMM. The planning module analyzes the human driver's goals or required assistance, identifies specific HAC goals, and organizes HAC tasks. Finally, the control module converts HAC tasks into control codes for cockpit devices to execute these tasks.

LangChain is an open-source framework designed to integrate LMMs with external APIs and data sources for broader application scenarios. HCopilot uses the LangChain framework to implement prompt templates and variable replacement in outputs, call tools API, facilitate information transfer between modules, and automate context summarization.

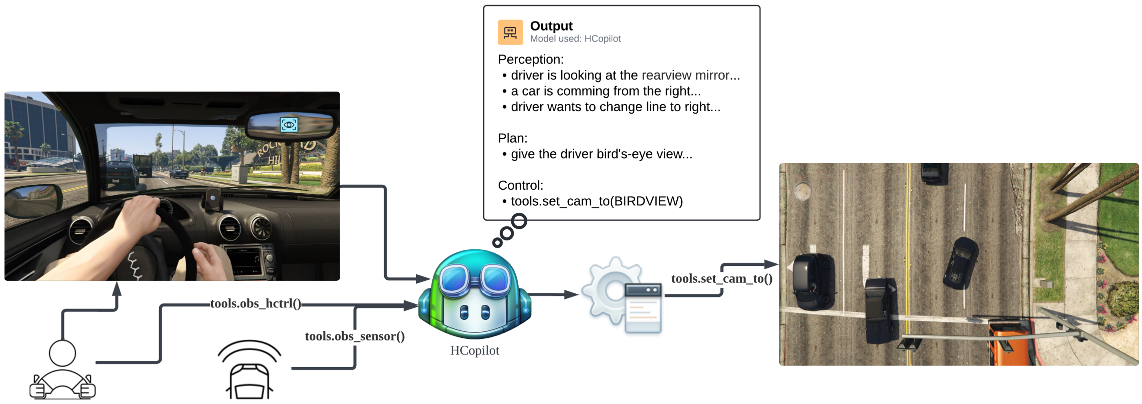

Figure 5: Scenario 1: Lane Change

Chapter 4: Results and Discussion

This section demonstrates HCopilot's capabilities in real urban and traffic scenarios within GTAV for situation awareness, task orchestration, and control, enhancing the driver's experience and safety through five designed city driving scenarios. The prototype HCopilot effectively enhances situational awareness through HUD and voice alerts based on the human driver's behavior and external environment changes, thus improving driving experience and safety.

4.1 Scenario 1: Lane Change

As shown in Figure 5, the human driver attempts a right lane change. HCopilot, observing the driver's eye movement and control data, accurately assesses the driver's intent and opts to display a real-time bird's-eye view on the HUD to enhance situational awareness, aiding the driver in safely completing the lane change. Even after the driver notices the rear vehicles, HCopilot continues to display the video stream, considering the enhancement of the driving experience.

4.2 Scenario 2: Turn Right at Intersection

In this scenario, as shown in Figure 6, the human driver is turning right at an intersection, focusing on vehicles in the opposite lane. HCopilot accurately judges the driver's intent and realizes the driver has not noticed traffic from the sides, thus choosing to display a bird's-eye view on the HUD and issuing a voice alert.

When reaching the moment shown in Figure 7, the driver's gaze indicates awareness of the oncoming traffic from the left, prompting HCopilot to cease the video stream, showcasing its intelligence in assisting the driver.

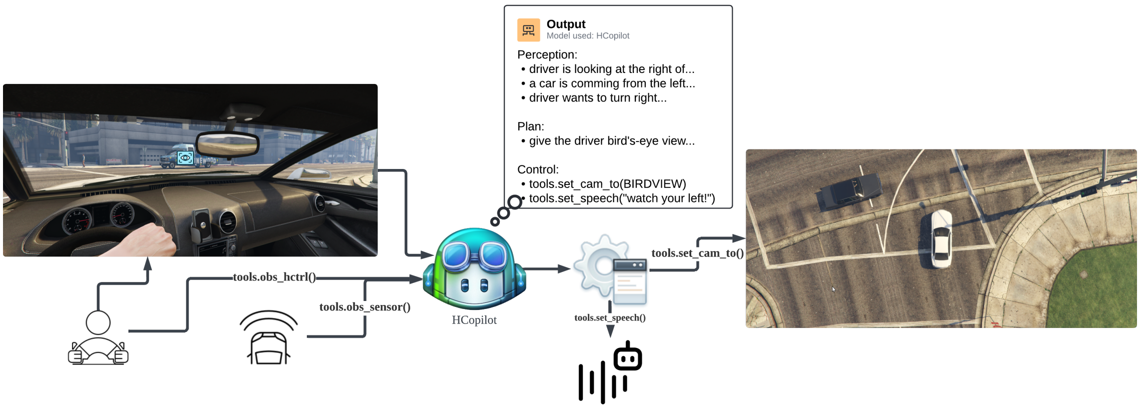

Figure 6: Scenario 2.1: Turn Right at Intersection

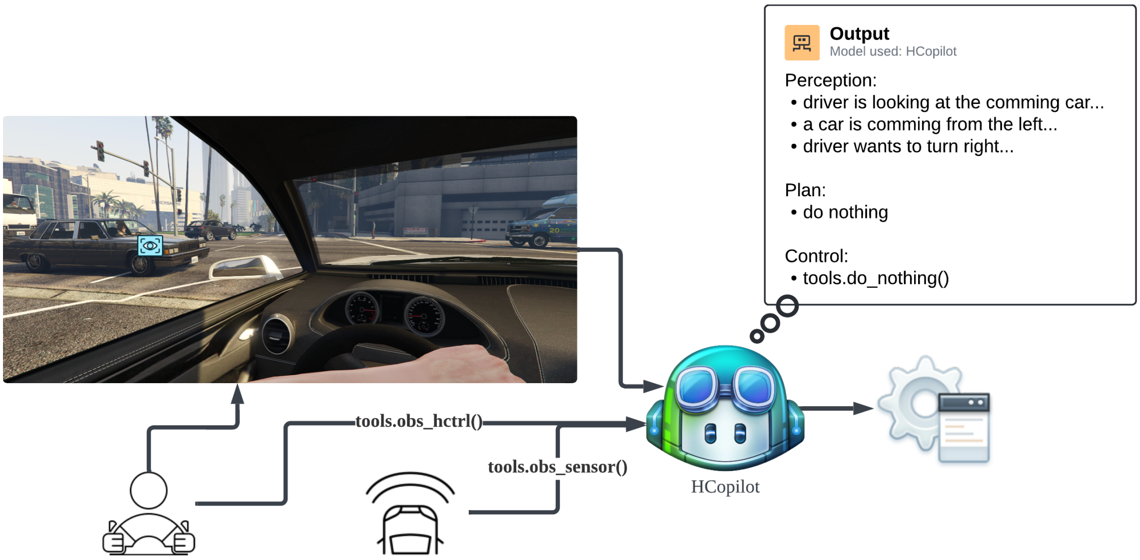

Figure 7: Scenario 2.2: Turn Right at Intersection

4.3 Scenario 3: Too Bright to See

In the scenario depicted in Figure 8, the driver encounters direct sunlight. A vehicle from the left approaches, unnoticed by the driver due to glare, who opts to proceed slowly. HCopilot, noting this, chooses to display a real-time video of vehicle v-100 on the HUD and issues a voice alert, prioritizing the threat over enhancing spatial awareness.

4.4 Scenario 4: Blind Area

In the scenario shown in Figure 9, the driver is checking a blind spot. HCopilot opts to enhance situational awareness by displaying a real-time video of the vehicle in the blind spot on the HUD, without using a voice alert as the driver is already aware of the potential threat.

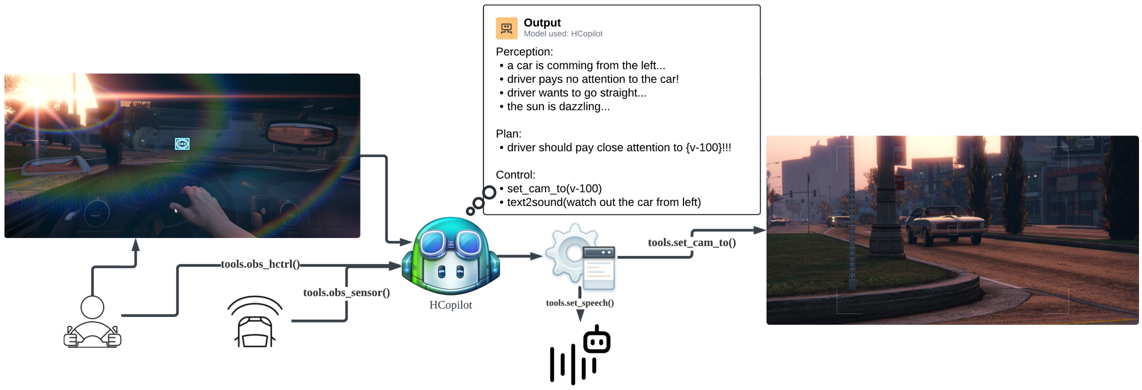

Figure 8: Scenario 3: Too Bright to See

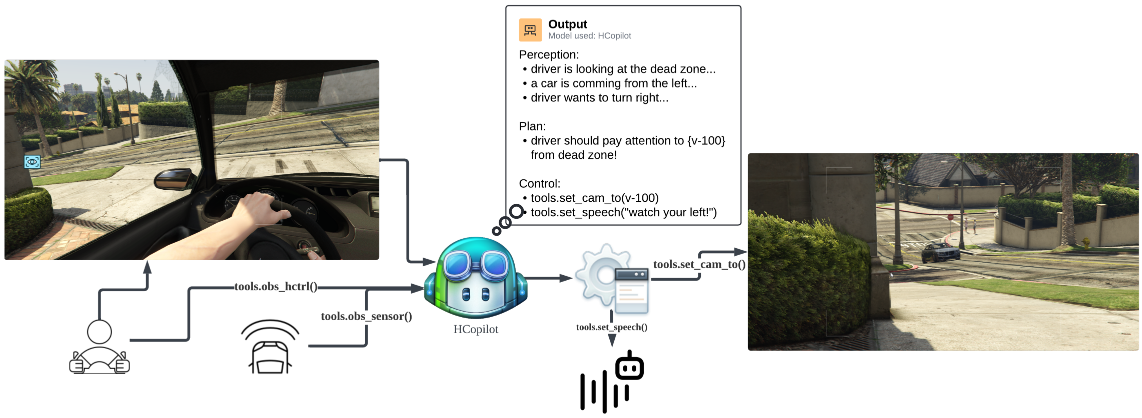

Figure 9: Scenario 4: Blind Area

4.5 Scenario 5: Low Attention

In the scenario shown in Figure 10, the driver is approaching an intersection, focused on a mobile phone. HCopilot chooses to issue a voice alert for the driver to prepare to brake, to avoid a collision with the vehicle ahead.

4.6 Discussion

HCopilot's implementation showcases its effectiveness and adaptability in complex urban driving environments. Through five different driving scenarios, it is observed how HCopilot enhances the driver's situational awareness and safety through real-time video streams and voice alerts. Specifically:

1. **Effectiveness in Enhancing Situational Awareness**: In all scenarios, HCopilot effectively enhances environmental awareness by providing additional visual information and voice alerts, aiding drivers in making safer decisions, especially during lane changes, blind spot checks, and low attention scenarios.

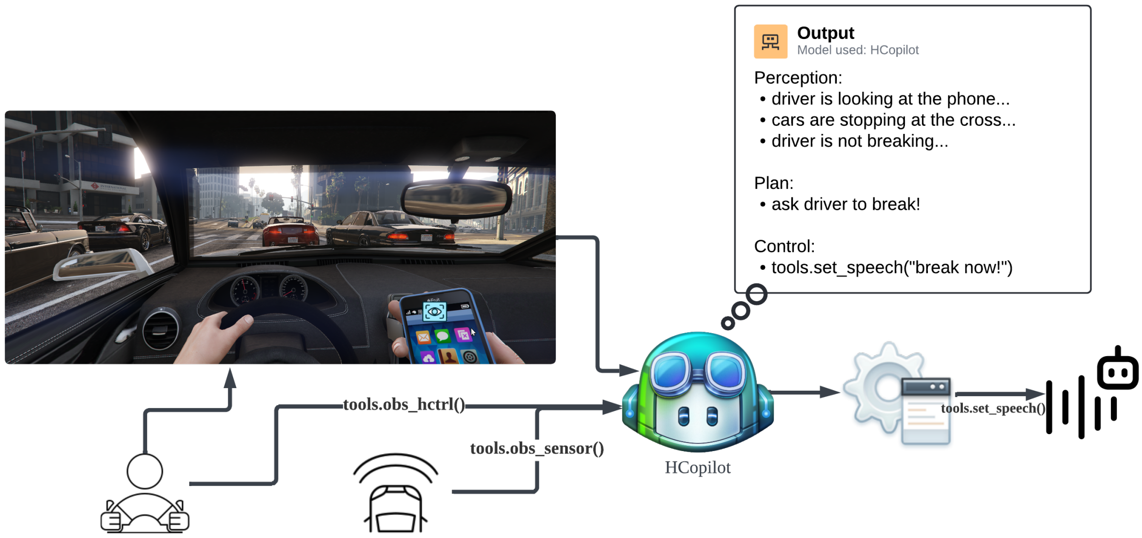

Figure 10: Scenario 5: Low Attention

2. **Adaptability and Intelligence of the System**: HCopilot demonstrates adaptability by adjusting its assistance strategy based on the driver's behavior and changes in the external environment. For example, it ceases video streaming when it detects that the driver has noticed a potential threat, reducing information overload and showing an understanding of human behavior.

3. **Enhancement of Driving Experience**: While HCopilot's interventions primarily enhance safety, they also improve the driving experience by reducing uncertainty and stress, making the driving process more relaxed and enjoyable. However, this positive impact needs validation across a broader driving population to ensure the system's universal applicability and acceptance.

Chapter 5: Conclusion and Further Work

5.1 Conclusion

This paper introduces a novel human-agent collaboration (HAC) framework centered around large multimodal models (LMMs) to enhance situational awareness (SA) in the cockpit. By integrating the perceptual and control capabilities of human drivers and AI driving systems, the HCockpit architecture successfully facilitates effective communication and collaboration between humans and machines. The experimental results demonstrate significant improvements in the driving experience and safety, especially in scenarios involving blind spots and distracted driving.

The technical implementation leverages the latest LMM, such as GPT-4 Turbo, allowing HCockpit to understand complex traffic environments and human driver intentions, generating appropriate tasks and control commands. Additionally, the designed communication mechanism ensures transparency in system decisions, helping to enhance human understanding and trust in the system.

Despite achievements, challenges such as precise alignment of multimodal data and real-time data processing to meet system performance requirements were encountered. Solutions to these issues largely rely on continuous optimization of data processing techniques and algorithm improvements.

Given more time for this project, exploration of advanced techniques for multimodal data fusion to improve system accuracy and response times would be pursued. Further, incorporating more user feedback mechanisms to optimize the human-machine interaction interface would be considered.

5.2 Reflection

This project was not only a technical challenge but also a profound learning and growth experience. Through this research, a deeper understanding of the potential and challenges of artificial intelligence in practical applications was gained. Moreover, experiences in team discussions taught effective communication and problem-solving under pressure, significantly benefiting professional development.

The successful implementation of the project validates the research direction and methodology, positively impacting academic and personal development. It enhanced technical capabilities and improved project management and decision-making skills.

5.3 Future Work

Future research will continue in the following directions:

1. **Entity Alignment Optimization**: Improving entity alignment techniques between different perspectives to enhance overall system performance and accuracy.
2. **Gaze Point Embedding Techniques**: Exploring embedding driver gaze data in vector form into the system decision-making process for more accurate behavior prediction and assistance.
3. **Local Real-Time Operation of LMMs**: Researching how to run large multimodal models in real-time locally, reducing reliance on cloud computing resources and enhancing system responsiveness and reliability.
4. **Unified and Standardized Cockpit Device Control Interface**: Collaborating with industry peers to promote the establishment of a unified and standardized cockpit device control interface, enabling seamless integration from high-level HAC tasks to specific device controls.

Through these studies, further enhancement of HCockpit's performance and practicality, contributing to the development of future intelligent transportation systems, is anticipated.

References

[1] D'Anielloa G, Loiab V, Orciuolib F. Ambient Systems , Networks and Technologies (ANT 2017) Adaptive Goal Selection for Improving Situation Awareness : The Fleet Management Case Study; 2018. .

[2] Yang Y, Zhang Q, Li C, Marta DS, Batool N, Folkesson J. Human-Centric Autonomous Systems with LLMs for User Command Reasoning. ArXiv. 2023;abs/2311.08206.

[3] Münst W. Prediction of Driver Behavior and Decision Strategies for Autonomous Driving: Using Machine Learning and Decision Theory. MyCoRe Community; 2020. .

[4] Gong X, Yingxue Y, Liu Y, Gong Q. Interaction Design of Closed Dark Cabin Driving Interface Based on Situation Awareness. 2023.

[5] Chen Q, Li J, Tei K. Attention-Guiding Takeover Requests for Situation Awareness in Semi-autonomous Driving. Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. 2023 Mar:416-21.

[6] de Winter JCF, Petermeijer SM, Abbink DA. Shared Control versus Traded Control in Driving: A Debate around Automation Pitfalls. Ergonomics. 2023 Oct;66(10):1494-520.

[7] Liao H, Shen H, Li Z, Wang C, Li G, Bie Y, et al.. GPT-4 Enhanced Multimodal Grounding for Autonomous Driving: Leveraging Cross-Modal Attention with Large Language Models. arXiv; 2023.

[8] Daronnat S, Azzopardi L, Halvey M. Comparing Levels and Types of Situational-Awareness Based Agent Transparency in Human-Agent Collaboration. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2022 Sep;66(1):1169-73.

Acknowledgment

I would like to express my deepest gratitude to everyone who contributed to the success of this project. First and foremost, I extend my sincere thanks to my academic advisor, whose invaluable guidance, patience, and expert advice were crucial throughout this research. His/Her mentorship was not only a professional anchor but also a great personal inspiration.

Special thanks go to my colleagues and the research team at BUPT and QMUL. Their insights and critiques have been immensely helpful and have greatly enriched my work. I am particularly grateful to them, whose collaboration and support have been indispensable.

I am also thankful to the technical staff and administrative personnel at University for their assistance in facilitating the resources needed for my experiments and research activities.

My appreciation also extends to the participants and volunteers who willingly engaged with our prototypes and provided feedback that was essential to refining our systems.

Finally, I must express my profound gratitude to my family and friends, who have provided me with moral support and encouragement throughout my academic journey. Their belief in my work and their constant encouragement have been a source of motivation and resilience.

Thank you all for your invaluable contributions to this project.

Appendices

Disclaimer

This report is submitted as part requirement for the undergraduate degree programme at Queen Mary University of London, and Beijing University of Posts and Telecommunications. It is the product of my own labour except where indicated in the text. The report may be freely copied and distributed provided the source is acknowledged.

BUPT No.: 2020213350

QM No.: 200980434

Full Name (Pin Yin): Hu Yitong

Full Name (Chinese): 胡逸同

Signature: 胡逸同

Date: 2024-04-23

Project specification

北京邮电大学 本科毕业设计（论文）任务书

Project Specification Form

Part 1 - Supervisor

<table>
<tr><td>论文题目
Project Title</td><td colspan="3">Design and development of a human-agent collaboration model for situation-awareness in cockpit</td></tr>
<tr><td>题目分类
Scope</td><td>Software Development</td><td>Implementation</td><td>Software and Hardware</td></tr>
<tr><td>主要内容
Project description</td><td colspan="3">To fulfil this project, you are expected to familiarize yourself first with the field of human-computer interaction (HCI) and the related specific areas, such as human-agent collaboration, situation awareness, context awareness, implicit interaction, natural interaction, HCI in cockpit etc.. You are then expected to do requirement analysis through user-centered methods and come up with the pain points and key requirements of human-agent collaboration in cockpit. Grounded in requirements, you need design and develop a system to solve the pain points uncovered previously and simulate or implement a system.Students applying for this project are required to design a software to support the collaboration between human and agents in intelligent vehicles. The software is expected to be user-centered and user-friendly to ensure the quality of experience of human drivers.</td></tr>
<tr><td>关键词
Keywords</td><td colspan="3">human-agent collaboration, context awareness, situation awareness, implicit interaction</td></tr>
<tr><td rowspan="4">主要任务
Main tasks</td><td colspan="3">1 Literature review on human-agent collaboration, situation awareness, context awareness, implicit interaction, natural interaction, HCI in cockpit</td></tr>
<tr><td colspan="3">2 Pain points and requirement analysis</td></tr>
<tr><td colspan="3">3 Design and develop a system with intelligent agent to facilitate the teaming work of human and agent</td></tr>
<tr><td colspan="3">4 Conduct experiment to evaluate the system</td></tr>
<tr><td rowspan="3">主要成果
Measurable outcomes</td><td colspan="3">1 Feature list of the software
supporting human-agent collaboration in intelligent vehicles</td></tr>
<tr><td colspan="3">2 An interactive machine-learning model supporting human-agent collaboration</td></tr>
<tr><td colspan="3">3 A software to demonstrate the performance of the interactive machine-learning model</td></tr>
</table>

北京邮电大学 本科毕业设计（论文）任务书

Project Specification Form

Part 2 - Student

学院 **School**	International School	**专业** **Programme**	**e-Commerce Engineering with Law**		
姓 **Family name**	Hu	**名** **First Name**	Yitong		
BUPT 学号 **BUPT number**	2020213350	**QM 学号** **QM number**	200980434	**班级** **Class**	2020215111
论文题目 **Project Title**	Design and development of a human-agent collaboration model for situation awareness in cockpit				
论文概述 **Project outline** **Write about 500-800 words** **Please refer to Project Student Handbook section 3.2**	(see below)				

Introduction

The rapid advancements in technology have led to new possibilities in the automotive industry, particularly in the development of smart cockpits where human-agent collaboration plays a pivotal role. This project proposes the design and development of a human-agent collaboration model to bolster situation awareness within a cockpit environment for SAE level 2 tasks. It will be grounded in Human-Computer Interaction (HCI) principles, focusing on aspects such as context awareness and implicit interaction to ensure natural and efficacious human-agent teaming.

User Requirements Analysis

Understanding the needs of drivers is crucial for the design of an effective human-agent collaboration model. We will adopt a user-centered approach to precisely capture the requirements. This process will commence with ethnographic studies and contextual research within the cockpit environment. A critical task analysis will also be facilitated to pinpoint the decision-making pain points during critical stages of vehicle operation.

Algorithms, Methodologies, and Techniques

The system will leverage various algorithms to ensure efficient human-agent collaboration, including models for recognizing driver's intentions and decision-making (how to help driver), and natural language processing for effective communication between the human and the agent.

I will employ adaptive modeling techniques to tailor interaction dynamically based on user physical state and context. A key focus will be on designing an algorithm capable of interpreting implicit cues from the human operator, such as gaze direction or control input patterns, to

	predict intentions and work as a copilot to adjust system behavior proactively. For situation awareness, multi-sensor data fusion will be essential, integrating inputs from radar, telemetry, environmental sensors, and other available data sources to provide a comprehensive operational picture to the agent for dynamical interaction. *Note: My preliminary plan is to obtain the above-mentioned data through a simulation platform.* I will also implement eye-tracking and biometric monitoring through camera as part of the user-system interaction studies. **User Interaction with the System** The interaction framework will emphasize intuitiveness and ease of use. It will include: • Voice commands or implicit cues (e.g., gaze) from user, and multimodal feedback (visual, auditory) for communication from system. *Note: Since the software is only used to demonstrate system performance, the above features will be simplified, and may be changed with consideration of the insights gained from requirements analysis.* **Experiments** To validate my hypotheses concerning enhanced situation awareness and collaboration efficacy, I will conduct a series of experiments: • Simulator-based testing or benchmark datasets to assess system responsiveness and suitability in diverse contextual scenarios. • Usability testing to evaluate the user experience and identify areas for refinement. **Tools, Languages, and Hardware** The software will be developed in Python, given its robust machine learning and data processing libraries and wide acceptance in research communities. For database management, SQL liked product will be utilized due to its reliability and performance with complex queries. The choice of hardware will depend on the processing needs identified during the initial prototyping stage but is anticipated to include cameras and microphones to collect data, as well as a computing platform with GPU acceleration.

	Background Material The following is a list of the background material that will guide the initial phase of the research: • M. Endsley, "Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors Journal 37(1), 32-64," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, pp. 32–64, Mar. 1995, doi: 10.1518/001872095779049543. • D. A. Norman, *The Design of Everyday Things*. USA: Basic Books, Inc., 2002. • B. Reeves and C. Nass, "The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Pla," *Bibliovault OAI Repository, the University of Chicago Press*, Jan. 1996. • Y. Xing *et al.*, "Driver Lane Change Intention Inference for Intelligent Vehicles: Framework, Survey, and Challenges," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4377–4390, May 2019, doi: 10.1109/TVT.2019.2903299. • Online HCI resources such as Interaction Design Foundation, Nielsen Norman Group, and ACM Digital Library. • Current standards and guidelines from the Federal Aviation Administration (FAA) and National Highway Traffic Safety Administration (NHTSA). • HCI in automotive contexts from journals like Automotive UI. *Note: These materials are initial guides, and the list is expected to expand during the literature review stage.*
道德规范 **Ethics**	Please confirm by checking the box: ☒ I confirm that I have discussed ethical issues with my supervisor.

Please discuss ethical issues with your supervisor using the ethics checklist in Project Handbook Appendix 1.	Summary of ethical issues: (write “None” if no ethical issues) None.
中期目标 **Mid-term target.** **It must be tangible outcomes, E.g. software, hardware or simulation.** **It will be assessed at the mid-term oral.**	By February 28, the goal would be to have the design of the intelligent agent system significantly underway with a basic prototype developed. The output by this time should include documentation of the requirements and a well-understood literature background from which to further develop the project. Machine learning integration should be in its initial stages of development, with consideration of the insights gained from the literature review and requirements analysis.

Work Plan (Gantt Chart)

Fill in the sub-tasks and insert a letter X in the cells to show the extent of each task

	Nov 1-15	Nov 16-30	Dec 1-15	Dec 16-31	Jan 1-15	Jan 16-31	Feb 1-15	Feb 16-28	Mar 1-15	Mar 16-31	Apr 1-15	Apr 16-30
Task 1 Literature review on human-agent collaboration, situation awareness, context awareness, implicit interaction, natural interaction, HCI in cockpit												
1.1 Identify Key Sources	X	X										
1.2 Analyse HCI Principles in Cockpit Context		X										
1.3 Synthesize Implicit & Natural Interaction Insights		X	X									
1.4 Document Review Insights			X									
Task 2 Pain points and requirement analysis												
2.1 Capture the Requirements through Literature Survey			X	X								
2.2 Requirements Documentation				X	X							
Task 3 Design and develop a system with intelligent agent to facilitate the teaming work of human and agent												
3.1 System Architecture Planning					X	X						
3.2 Prototype Development						X	X					
3.3 Integrate Machine Learning Model							X	X				
3.4 Iterative Design & Testing							X	X	X			
Task 4 Conduct experiment to evaluate the system												
4.1 Experiment Design									X	X		
4.2 Simulation & User Testing										X	X	
4.3 Data Collection & Analysis											X	X
4. Refinement & Documentation												X

Early-term progress report

北京邮电大学 本科毕业设计（论文）初期进度报告

Project Early-term Progress Report

<table>
<tr><td>学院
School</td><td>International School</td><td>专业
Programme</td><td colspan="3">e-Commerce Engineering with Law</td></tr>
<tr><td>姓
Family name</td><td>Hu</td><td>名
First Name</td><td colspan="3">Yitong</td></tr>
<tr><td>BUPT 学号
BUPT number</td><td>2020213350</td><td>QM 学号
QM number</td><td>200980434</td><td>班级
Class</td><td>2020215111</td></tr>
<tr><td>论文题目
Project Title</td><td colspan="5">Design and development of a human-agent collaboration model for situation awareness in cockpit</td></tr>
</table>

1. Introduction

The advent of fully autonomous driving is fast approaching. However, until self-driving systems can adeptly manage a diversity of situational challenges, environmental variables, and unforeseen circumstances, the journey toward complete autonomy will be evolutionary, marked by the necessity for human oversight (Gao et al., 2021; X. Li et al., 2023; Y. Wang et al., 2023). Concurrently, as autonomous driving technology evolves, an array of new challenges arises. Presently, numerous Original Equipment Manufacturers (OEMs) are adopting Level 2+ or Level 3 autonomous driving capabilities that permit drivers to temporarily relinquish control of specific driving functions, thereby harmonizing vehicular performance with cost-effectiveness (Mangal, 2021).

However, such systems may intermittently necessitate human re-engagement in vehicle operation, and conversely, drivers might require support from autonomous systems in particular scenarios (Wang et al., 2020; Y. Wang et al., 2023). Achieving fluid communication and collaboration between the driver and the autonomous system is paramount at this level of autonomous driving to enhance driving experience and safety (W. Li et al., 2023; X. Li et al., 2023; Z. Yang et al., 2023).

Consequently, the industry is in pursuit of a sophisticated, which is core topic of this project, the interactive system—one that fosters situational awareness through the amalgamation of insights from both the vehicle's interior and the external environment, acting as a reciprocal link between the autonomous system and the driver, and orchestrating actions to facilitate human-agent collaboration tasks in the context of awareness.

2. Related Works

2.1. Situation Awareness

2.1.1. Introduction

Situation awareness (SA) is the perception of elements in the environment, comprehension of their meaning, and the projection of their status in the near future. It involves being aware of the start and end of situations, as well as any active situation at any given time. SA in intelligent cockpit refers to the ability of an autonomous vehicle to perceive its environment, understand the significance of the perceived information, and predict future states of the environment.

SA is a critical aspect of autonomous driving, as it directly impacts the safety and efficiency of the vehicle's operation. One main issue in SA includes attentional tunneling, where operators focus on a single goal and lose awareness of the overall picture, and stimuli that may divert attention from important aspects, leading to erroneous decisions (D'Anielloa et al., 2018).

2.1.2. Representative Works

The development of research hotspots in this field has been focused on improving the SA of autonomous vehicles through various methods, including multimodal sensing, machine learning, and interface design.

Multimodal Sensing for SA: (J. Yang et al., 2023) demonstrated a multimodal sensing approach for objective SA monitoring in autonomous driving. The study used physiological sensor data from electroencephalogram and eye-tracking to assess SA. The results showed that a multi-physiological sensor-model outperformed the single sensing model, suggesting that multimodal sensing can objectively predict SA.

Machine Learning for SA: Research by (Münst, 2020) used supervised machine learning techniques for both reactive predictions (short-term) and motivation-based predictions (long-term) to predict the behavior of other traffic participants and decide what to do with these. The study showed that even simple prediction and decision algorithms can considerably improve the current status quo, although more advanced models increase complexity.

Interface Design for SA: (Gong et al., 2023) explored the design of the vehicle terminal interface in a closed dark cabin driving environment to improve the driver's perception of the environmental information outside the cabin and the ease of use of the interface. The study found that the design method effectively enhanced the driver's SA.

Attention-guiding Techniques for SA: (Chen et al., 2023) proposed to improve drivers' takeover performance by utilizing attention-guiding techniques when delivering the takeover request (TOR) in semi-autonomous driving. The preliminary experiment indicated that this method reduced drivers' collision rate and mental workload.

2.1.3. Challenges

Despite the advancements in SA for autonomous driving, several challenges remain:

1. **Interface Design**: The design quality of the vehicle terminal interface directly affects the driver's SA level during driving. Therefore, improving the driver's perception of the environmental information and the ease of use of the interface is a challenge (Gong et al., 2023).
2. **Human-Automation Interaction**: A major question in human-automation interaction is whether tasks should be traded or shared between human and automation. This dilemma may impact the design of automation systems (de Winter et al., 2023).
3. **Traffic Sign Detection**: Traffic sign detection and recognition is a critical aspect of the environmental awareness module of autonomous driving. Early traffic sign recognition methods were mostly based on color features, shape features, or multi-feature fusion (H. Li et al., 2023).
4. **Transparency of agents**: In the context of automated vehicles, the transparency and reliability of in-vehicle intelligent agents significantly impact driver perception, workload, and SA (Daronnat et al., 2022).

2.2. Human-Agent Collaboration

2.2.1. Introduction

Human-Agent Collaboration (HAC) in intelligent cockpit refers to the interaction between human drivers and autonomous driving systems. This field aims to enhance the safety, efficiency, and user experience of autonomous vehicles by leveraging the strengths of both human drivers and autonomous systems. The main issues in this field include path planning, perception of the dynamic world, decision-making, and communication between human drivers and autonomous systems (Agapito and Fallon, 2022; Khemchandani et al., 2023; Plebe et al., 2022).

2.2.2. Representative Works

Applications of Large-Scale Foundation Models for Autonomous Driving:

(Huang et al., 2024) investigates the application of large language models (LLMs) and foundation models in autonomous driving. The authors propose that these models can be used to reformulate autonomous driving by leveraging human knowledge, common sense, and reasoning. The models can be applied in various areas, including simulation, world model, data annotation, and planning or end-to-end solutions.

(Liao et al., 2023) introduces a sophisticated encoder-decoder framework, the Context-Aware Visual Grounding (CAVG) model, to address visual grounding in autonomous vehicles. The model integrates five core encoders with a Multimodal decoder, enabling it to capture contextual semantics and learn human emotional features. The model demonstrated high prediction accuracy and operational efficiency, even with limited training data.

In conclusion, multi-modal large-scale models, such as GPT-4, have been applied in autonomous driving to enhance human-agent collaboration. These models can process and interpret a range of cross-modal inputs, yielding a comprehensive understanding of the correlation between verbal commands and corresponding visual scenes. They can also learn human emotional features, which can be useful in understanding and responding to human drivers' intentions and emotions (Cui et al., 2023; Liao et al., 2023; L. Wang et al., 2023; Z. Yang et al., 2023).

Real Time Human Assisted Path Planning for Autonomous Agent using VR:

(Khemchandani et al., 2023) focuses on path planning, a critical aspect of autonomous driving. The researchers developed a virtual reality (VR) system to train defense personnel in path planning for various operations in remote areas. The system simulates real-world scenarios, including traffic light systems, AI car navigation algorithms, and rescue operations, providing a cost-effective and safe training environment.

Toward Policy Explanations for Multi-Agent Reinforcement Learning:

(Boggess et al., 2022) presents novel methods to generate policy explanations for multi-agent reinforcement learning (MARL), a technique used in autonomous driving. The authors developed methods to summarize agent cooperation and task sequence and to answer queries about agent behavior. The study found that these explanations improved user performance and satisfaction.

Distributed cognition for collaboration between human drivers and self-driving cars:

This paper proposes a collaboration mechanism based on the concept of distributed cognition. The authors suggest that intelligence lies not only in the individual entities (human or autonomous agent) but also in their interaction. The study uses a driving simulator to demonstrate the collaboration in action, showing how the human can communicate and interact with the agent in various ways with safe outcomes (Plebe et al., 2022).

2.2.3. Challenges

Despite the progress made in human-agent collaboration in autonomous driving, several challenges remain. These include accurately representing the mutual effects of vehicles and modeling dynamic traffic environments in mixed autonomy traffic, which includes both autonomous vehicles and human-driven vehicles (Liu et al., 2022). Another challenge is managing the risk that an agent's action could harm a friendly computer, which must be balanced against the losses that could occur if the agent does not act (Kott, 2023). Lastly, there is a need for more research on how to maintain human expertise and relevance in professional decision-making as automation increases (X. Li et al., 2023).

3. Conclusion

Addressing the challenges highlighted above, this project introduces HarmonyCockpit (HCockpit), a framework that integrates advanced multi-modal large-scale models to facilitate transparent human-agent collaboration (HAC) within the cockpit environment. HCockpit cultivates situational awareness by synthesizing information from both the cockpit's internal and external milieus and directs actions in concert with established cockpit functions to support HAC tasks grounded in situational cognizance.

To assess HCockpit's efficacy and derive insights, the HarmonyCopilot (HCopilot) was developed as an operational example of the HCockpit framework, utilizing cutting-edge multi-modal large-scale models alongside conventional intelligent cockpit designs. As a reciprocal link between the autonomous driving system and the driver, HCopilot strives to augment the driving experience and safety via an integrated human-vehicle interface.

Different from (Huang et al., 2023; Liao et al., 2023; L. Wang et al., 2023; S. Wang et al., 2023; Wen et al., 2023)s' works, this project research introduces the HCockpit framework and the HCopilot exemplar as pioneering contributions to autonomous driving technology, emphasizing AI-driven collaboration between humans and machines. It notably accentuates situational awareness and undertakes passive human-system interaction. Leveraging substantial multi-modal models, the initiative endeavors to enrich comprehension of both vehicular confines and the external environment, thereby enhancing the response capability and transparency of the autonomous system. Departing from conventional autonomous driving studies, it incorporates advanced functionalities like semantic comprehension, driver intent prediction, and bidirectional communication—innovations that position it at the forefront of the field.

HCockpit notably excels in personalizing user experience by proactively adapting to the driver's behaviors and preferences, enhancing trust and satisfaction with the system. In terms of safety, the model responds promptly to lapses in the driver's focus or when faced with challenging driving scenarios that surpass the autonomous system's capacity, proactively signaling the driver to assume control. This feature serves to avert potential accidents and bolsters overall driving safety. Such advancements highlight HCockpit's role in not only improving autonomous driving performance but also in offering a tailored driving experience with significant market potential.

4. References

Agapito, L., Fallon, M., 2022. 19th on Robots and Vision.

Boggess, K., Kraus, S., Feng, L., 2022. Toward Policy Explanations for Multi-Agent Reinforcement Learning. https://doi.org/10.48550/arXiv.2204.12568

Chen, Q., Li, J., Tei, K., 2023. Attention-guiding Takeover Requests for Situation Awareness in Semi-autonomous Driving. Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction 416–421. https://doi.org/10.1145/3568294.3580118

Cui, C., Ma, Y., Cao, X., Ye, W., Zhou, Y., Liang, K., Chen, J., Lu, J., Yang, Z., Liao, K.-D., Gao, T., Li, E., Tang, K., Cao, Z., Zhou, T., Liu, A., Yan, X., Mei, S., Cao, J., Wang, Z., Zheng, C., 2023. A Survey on Multimodal Large Language Models for Autonomous Driving.

D'Anielloa, G., Loiab, V., Orciuolib, F., 2018. Ambient Systems , Networks and Technologies (ANT 2017) Adaptive Goal Selection for improving Situation Awareness : the Fleet Management case study.

Daronnat, S., Azzopardi, L., Halvey, M., 2022. Comparing Levels and Types of Situational-Awareness based Agent Transparency in Human-Agent Collaboration. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 66, 1169–1173. https://doi.org/10.1177/1071181322661498

de Winter, J.C.F., Petermeijer, S.M., Abbink, D.A., 2023. Shared control versus traded control in driving: a debate around automation pitfalls. Ergonomics 66, 1494–1520. https://doi.org/10.1080/00140139.2022.2153175

Gao Z., Li W., Liang J., Pan H., Xu W., Shen M., 2021. Trust in automated vehicles. Adv Psychol Sci 29, 2172–2183. https://doi.org/10.3724/SP.J.1042.2021.02172

Gong, X., Yingxue, Y., Liu, Y., Gong, Q., 2023. Interaction Design of Closed Dark Cabin Driving Interface based on Situation Awareness. https://doi.org/10.54941/ahfe1003792

Huang, J., Jiang, P., Gautam, A., Saripalli, S., 2023. GPT-4V takes the wheel: Evaluating promise and challenges for pedestrian behavior prediction. ArXiv abs/2311.14786.

Huang, Y., Chen, Y., Li, Z., 2024. Applications of Large Scale Foundation Models for Autonomous Driving. https://doi.org/10.48550/arXiv.2311.12144

Khemchandani, V., Khan, M.A., Barkaa, M.U., Chandra, S., Wadalkar, N.M., 2023. Real Time Human Assisted Path Planning for Autonomous Agent using VR. 2023 2nd Edition of IEEE Delhi Section Flagship Conference (DELCON) 1–6. https://doi.org/10.1109/DELCON57910.2023.10127333

Kott, A., 2023. Autonomous Intelligent Cyber-defense Agent: Introduction and Overview. https://doi.org/10.48550/arXiv.2304.12408

Li, H., Ma, Y., Yu, J., Zhang, Z., 2023. Studies Advanced in Traffic Sign Detection in Autonomous Driving Scenarios. ACE 8, 501–506. https://doi.org/10.54254/2755-2721/8/20230261
Li, W., Cao, D., Tan, R., Shi, T., Gao, Z., Ma, J., Guo, G., Hu, H., Feng, J., Wang, L., 2023. Intelligent Cockpit for Intelligent Connected Vehicles: Definition, Taxonomy, Technology and Evaluation. IEEE Transactions on Intelligent Vehicles 1–14. https://doi.org/10.1109/TIV.2023.3339798
Li, X., Bai, Y., Cai, P., Wen, L., Fu, D., Zhang, B., Yang, X., Cai, X., Ma, T., Guo, J., Gao, X., Dou, M., Li, Y., Shi, B., Liu, Y., He, L., Qiao, Y., 2023. Towards Knowledge-driven Autonomous Driving.
Liao, H., Shen, H., Li, Z., Wang, C., Li, G., Bie, Y., Xu, C., 2023. GPT-4 Enhanced Multimodal Grounding for Autonomous Driving: Leveraging Cross-Modal Attention with Large Language Models. https://doi.org/10.48550/arXiv.2312.03543
Liu, Q., Li, X., Li, Z., Wu, J., Du, G., Gao, X., Yang, F., Yuan, S., 2022. Graph Reinforcement Learning Application to Co-operative Decision-Making in Mixed Autonomy Traffic: Framework, Survey, and Challenges. https://doi.org/10.48550/arXiv.2211.03005
Mangal, N., 2021. Automated Driving Requires Rethink of Human-Vehicl. APTIV.
Münst, W., 2020. Prediction of Driver Behavior and Decision Strategies for Autonomous Driving: Using Machine Learning and Decision Theory. MyCoRe Community. https://doi.org/10.18445/20201122-144857-0
Plebe, A., Rosati Papini, G.P., Cherubini, A., Da Lio, M., 2022. Distributed cognition for collaboration between human drivers and self-driving cars. Front Artif Intell 5, 910801. https://doi.org/10.3389/frai.2022.910801
Wang, J., Zhang, L., Huang, Y., Zhao, J., 2020. Safety of Autonomous Vehicles. Journal of Advanced Transportation 2020, e8867757. https://doi.org/10.1155/2020/8867757
Wang, L., Ren, Y., Jiang, H., Cai, P., Fu, D., Wang, T., Cui, Z., Yu, H., Wang, X., Zhou, H., Huang, H., Wang, Y., 2023. AccidentGPT: Accident Analysis and Prevention from V2X Environmental Perception with Multi-modal Large Model.
Wang, S., Zhu, Y., Li, Z., Wang, Y., Li, L., He, Z., 2023. ChatGPT as Your Vehicle Co-Pilot: An Initial Attempt. IEEE Trans. Intell. Veh. 1–17. https://doi.org/10.1109/TIV.2023.3325300
Wang, Y., Jiao, R., Lang, C., Zhan, S.S., Huang, C., Wang, Z., Yang, Z., Zhu, Q., 2023. Empowering autonomous driving with large language models: A safety perspective. ArXiv abs/2312.00812.
Wen, L., Yang, X., Fu, D., Wang, X., Cai, P., Li, X., Ma, T., Li, Y., Xu, L., Shang, D., Zhu, Z., Sun, S., Bai, Y., Cai, X., Dou, M., Hu, S., Shi, B., Qiao, Y., 2023. On the road with GPT-4V(ision): Early explorations of visual-language model on autonomous driving. ArXiv abs/2311.05332.
Yang, J., Liang, N., Pitts, B.J., Prakah-Asante, K.O., Curry, R., Blommer, M., Swaminathan, R., Yu, D., 2023. Multimodal Sensing and Computational Intelligence for Situation Awareness Classification in Autonomous Driving. IEEE Trans. Human-Mach. Syst. 53, 270–281. https://doi.org/10.1109/THMS.2023.3234429
Yang, Z., Jia, X., Li, H., Yan, J., 2023. LLM4Drive: A Survey of Large Language Models for Autonomous Driving. arXiv.org.

是否符合进度？On schedule as per GANTT chart? YES.
下一步 Next steps: 3.2 Prototype Development 3.3 Integrate Machine Learning Model

Mid-term progress report

北京邮电大学 本科毕业设计（论文）中期进度报告

Project Mid-term Progress Report

学院 School	International School	**专业 Programme**	**e-Commerce Engineering with Law**		
姓 Family name	Hu	**名 First Name**	Yitong		
BUPT 学号 BUPT number	2020213350	**QM 学号 QM number**	200980434	**班级 Class**	2020215111
论文题目 Project Title	Design and development of a human-agent collaboration model for situation awareness in cockpit				

是否完成任务书中所定的中期目标？Targets met (as set in the Specification)?
YES.

已完成工作 Finished work:

1. Literature Review and Requirement Analysis

The advent of fully autonomous driving is fast approaching. However, until self-driving systems can adeptly manage a diversity of situational challenges, environmental variables, and unforeseen circumstances, the journey toward complete autonomy will be evolutionary, marked by the necessity for human oversight. Concurrently, as autonomous driving technology evolves, an array of new challenges arises. Presently, numerous Original Equipment Manufacturers (OEMs) are adopting Level 2+ or Level 3 autonomous driving capabilities that permit drivers to temporarily relinquish control of specific driving functions, thereby harmonizing vehicular performance with cost-effectiveness.

However, such systems may intermittently necessitate human re-engagement in vehicle operation, and conversely, drivers might require support from autonomous systems in particular scenarios. Achieving fluid communication and collaboration between the driver and the autonomous system is paramount at this level of autonomous driving to enhance driving experience and safety.

Consequently, the industry is in pursuit of a sophisticated, which is core topic of this project, the interactive system—one that fosters situational awareness through the amalgamation of insights from both the vehicle's interior and the external environment, acting as a reciprocal link between the autonomous system and the driver, and orchestrating actions to facilitate human-agent collaboration tasks in the context of awareness.

All in all, the demand for human-agent cooperation in smart cockpit is mainly reflected in improving the user experience and safety of driving:

1. Improve user experience (UX):
    1. When the driver attempts to change lanes, reverse or turn around, the corresponding image will be displayed or a voice/text reminder will be initiated according to the cockpit function

2. When the driver initiates active interaction, the corresponding functions of the cockpit are mobilized to meet user needs.
3. Generate suitable autonomous driving preference suggestions based on driver status/habits, for example: when the driver drinks water, give priority to a smooth driving strategy

2. Improve security:
    1. When the driver's attention level is low or unable to respond to a crisis, a reminder action is generated or the autonomous driving system is recommended to take over control.
    2. When the driver is not aware of potential threats around him, he proactively displays the source of the threat
    3. When encountering complex road conditions that the autonomous driving system is not capable of handling, assess whether the driver is qualified to take over, and then ask the driver to take over the vehicle or remind the driver to pay more attention and prepare to take over.

2. System Architecture Planning

2.1 Introduction

This project introduces the HarmonyCockpit (HCockpit), a framework designed to incorporate cutting-edge multi-modal large models and orchestrate human-agent collaboration (HAC) in cockpit with transparency.

The HCockpit achieves situational awareness by integrating insights into both the internal and external environments of the cockpit and generates natural language level tasks in context-aware to facilitate HAC. Subsequently, HCockpit translates tasks into actions adapting to predefined cockpit functions, thereby guiding both humans and agents towards coordinated action. Metaphorically speaking, within a cockpit setting that includes a human driver and an autonomous driving system (agent), the HCockpit serves as the orchestrating force between human and agent.

To assess the efficacy of the HCockpit and garner deeper understandings, this project developed HarmonyCopilot (HCopilot) as an instance of the HCockpit framework by leveraging state-of-the-art multi-modal large models and incorporating advanced smart cockpit models.

2.2 Preliminaries

This project defines actions of the environment and agents as follows. The global environment is defined as consisting of both the external environment outside the ego vehicle and the internal cockpit environment. Within the global environment, there are three types of agents: the autonomous driving system (AI driver), the human driver, and the HCopilot. The first two are considered independent homomorphic agents, meaning they have the same observation space (the external environment) and action space (vehicle control); differing from the first two, the HCopilot: 1. has global observation, 2. its output actions control cockpit devices. The core of HCopilot is driven by language models, and its strategy is interpretable and auditable to ensure driving safety.

2.3 System Architecture Design

Input: Comprehensive situational awareness towards the global environment

HCopilot needs to continuously observe the global environment (outside + inside the cockpit) to achieve situational awareness (SA). For the internal cockpit environment, this project only considers the main driver inside the cockpit. HCopilot receives multimodal data inputs including driver's natural language commands, physiological and control data. It utilizes visual-language large models to extract semantic information of the scene and the driver's intention, then saves it as a structured state vector for observation of the human driver.

For the external environment, this project considers that HCopilot should obtain observations of the external environment through the AI driver, and therefore selects a visual-language large model fine-tuned in driving scenarios as simulator for the AI driver. This simulator provides interpretable environmental semantic information and projections of future states and events according to the vehicle's external environment. Similarly, HCopilot saves the aforementioned information as a state vector for observation of the external environment.

It should be noted that HCopilot does not observe another agent—the state of the AI driver, because when the human driver is in control of the vehicle, there is no need to consider the AI driver's strategy or intention; conversely, when the AI driver controls the vehicle, the human driver's intentions or preferences become crucial.

Decision: orchestrate HAC tasks in context-aware

HCopilot combines comprehensive SA of the global environment and context, serving as a messenger and coordinator for the driver and AI driver, orchestrating actions to facilitate HAC, e.g., controlling relevant hardware inside the cockpit to send alerts or information to the driver and providing control suggestions or takeover requests to the AI driver according to the driver's situation. In this process, context awareness is essential. Inspired by the work of Google RT-2, this project designed a memory module for HCockpit, enabling HCopilot to provide more comprehensive task orchestrations based on historical SA.

It is worth noting that there is still a considerable gap between generating HAC tasks and controlling specific cockpit devices because it involves details of the cockpit device control interface, which is beyond the scope of this project's research. HCockpit focuses on the generation and orchestration of more general, high-level HAC tasks, so it designed a simple device control interface for evaluation and demonstration. HCopilot is an instance that has adopted the aforementioned cockpit model.

3. Integrating Machine Learning Model

The HCopilot system incorporates advanced machine learning models to handle various tasks essential for enhancing human-agent collaboration (HAC) in a smart cockpit environment. This section elaborates on the integration of specific models like Pathways Language and Image model (PaLI-X), Pathways Language model Embodied (PaLM-E), GPT-4 variants, and OpenAI Codex for different functionalities within the HCopilot framework. However, integrating ML models is still undergoing and only a part of codes will be demonstrated.

3.1 Observation and Semantic Interpretation

Model: PaLI-X and PaLM-E

Functionality: These models are crucial for interpreting the multimodal data collected from both the internal and external environments of the cockpit. PaLI-X, with its prowess in language and image understanding, processes the visual inputs and natural language communications from the driver to understand the current situation inside and outside the cockpit. PaLM-E, being an embodied model, further enhances the system's capability to predict and propose actions based on the interpreted data, adding a layer of "understanding" to the system's decision-making process.

Implementation:

```
from pali_x import PaLIX
from palm_e import PaLME

# Initialize models
pali_x_model = PaLIX()
palm_e_model = PaLME()

def observe_and_interpret(internal_data, external_data):
    # Observing Internal Environment
    internal_state_vector = pali_x_model.process_internal_data(internal_
data)

    # Observing External Environment through PaLIX
    external_state_vector = pali_x_model.process_external_data(external_
data)

    # Using PaLME for action recommendation based on observed states
    action_suggestions = palm_e_model.recommend_actions(internal_state_v
ector, external_state_vector)

    return action_suggestions
```

3.2 Situational Awareness through Multi-Prompt-Tuning GPT-4 Variants

Model: GPT-4-V (Multiple instances with prompt-tuning for specific observational tasks)

Functionality: Customized GPT-4-V instances, each tuned with specific prompts to handle distinct situational awareness tasks. These instances analyze the state vectors provided by PaLI-X to generate natural language tasks that are context-aware, aiding in precise HAC task orchestration.

Implementation:

```
from gpt4v_instance import GPT4V

# Initialize GPT-4-V Instances for Different Observational Contexts
gpt4v_driving_conditions = GPT4V(prompt="Analyze driving conditions:")
gpt4v_driver_status = GPT4V(prompt="Evaluate driver's state:")

def generate_context_aware_tasks(internal_state_vector, external_state_v
ector):
    # Situational Analysis for Driving Conditions
    driving_condition_tasks = gpt4v_driving_conditions.generate_tasks(ex
ternal_state_vector)

    # Driver's Status Evaluation
    driver_status_tasks = gpt4v_driver_status.generate_tasks(internal_st
ate_vector)

    combined_tasks = {"driving_conditions": driving_condition_tasks, "dr
iver_status": driver_status_tasks}
    return combined_tasks
```

3.3 Translating HAC Tasks to Cockpit Actions

Model: OpenAI Codex

Functionality: Utilizes OpenAI Codex to interpret natural language HAC tasks generated by the GPT-4 variants and translate them into specific, executable commands for the cockpit's hardware and software interface, ensuring the HCockpit's integration with the vehicle's operational systems is seamless and efficient.

Implementation:

```
from openai_codex import OpenAICodex

codex = OpenAICodex()

def execute_cockpit_tasks(tasks):
    for task in tasks:
        # Translate HAC tasks to cockpit function API calls
        code_command = codex.translate_to_code(task)
        exec(code_command)  # Execute the translated command in a safe a
nd controlled environment
```

```
# Example Cockpit Functions API (Pseudocode)
def adjust_seat(position):
    # API Call to Adjust the Seat to Desired Position
    pass

def display_alert(message):
    # API Call to Display an Alert on the Cockpit's Screen
    pass
```

尚需完成的任务 Work to do:

For the HCockpit project, iterative developing and testing are critical steps in developing a robust system ready for real-world applications. Conducting experiments using simulators like GTA V can offer valuable insights into the system's performance under various scenarios, which closely mimic real-life conditions without the associated risks and costs.

1. Conduct Experiments to Evaluate the System

Using Simulators: Choose simulators like GTA V for its realistic urban environments, sophisticated physics engine, and dynamic weather conditions. These features provide a diverse range of driving scenarios, including pedestrian interactions, traffic congestion, and emergency situations, ideal for testing the HCockpit system.

Integration: Develop an integration layer that allows the HCockpit to communicate with the simulator. This layer translates the simulator's sensory data into inputs the HCockpit can understand and vice versa.

Evaluation Metrics: Define key performance indicators (KPIs) such as response time to external events, accuracy in executing given tasks, ease of human-agent collaboration, and system reliability under different conditions.

2. Develop a Demo

Scenario Planning: Create a variety of scenarios within the simulator that highlight the HCockpit's capabilities, such as emergency takeover, assisting human to turn around.

Interactivity: Ensure the demo allows for real-time interaction, giving users the ability to issue commands or change parameters and observe the system's response.

Visualization: Use visualization tools to clearly demonstrate the HCockpit's situational awareness, decision-making process, and action execution. This could include visual overlays on the simulator's output showing detected objects, planned paths, and decision points.

存在问题 **Problems:** 1. **Real-Time Inference and Deployment Cost:** Large model inference suffers from latency issues, and deploying such models locally incurs significant costs due to the need for powerful computing resources. 2. **Sampling Rate for Human and External Environment Data:** Determining the appropriate sampling rate for human data (e.g., physiological and behavioral indicators) and external environmental data is crucial for maintaining system responsiveness and accuracy without overwhelming the system with excessive data. 3. **Integration with Existing Vehicle Systems:** Integrating the HCockpit system with existing vehicle systems could pose compatibility and operational challenges, as vehicles have diverse hardware and software configurations. 4. **User Acceptance and Trust:** Convincing drivers to trust and accept the AI's decision-making, especially in critical and emergency scenarios, remains a considerable challenge.
拟采取的办法 **Solutions:** 1. **Model Optimization:** – Invest in optimizing the AI models for efficiency without significantly compromising accuracy, using techniques such as quantization, pruning, and knowledge distillation. 2. **Adaptive Sampling Rate:** – Implement an adaptive sampling algorithm that adjusts the rate based on the current driving scenario's complexity. For instance, increase the sampling rate in high-risk situations such as urban driving or bad weather conditions, and decrease it during steady highway cruising. – Use active learning strategies to identify and prioritize the most informative data points, optimizing the trade-off between system performance and computational load. 3. **Modular Integration Framework:** – CodeX is designed to be integrated for generating actions to control cockpit according to the predefined cockpit profile. – The seamless connection of HAC tasks to cockpit equipment control requires a unified and standardized interface, which requires the joint efforts of colleagues in the industry. 4. **Iterative Trust-Building:**

– Deploy the system initially in less critical functions, allowing users to experience and understand the technology in low-risk environments.

– Provide detailed feedback and explanations for the AI's decisions, especially when those decisions override or suggest against the driver's actions. That's one of the main reason LLM is used as core in HCockpit.

论文结构 Structure of the final report: (Chapter headings and section sub headings)

Table of Contents

Supervision log

北京邮电大学 本科毕业设计（论文）教师指导记录表

Project Supervision Log

<table>
<tr><td>学院
School</td><td>International School</td><td>专业
Programme</td><td colspan="3">e-Commerce Engineering with Law</td></tr>
<tr><td>姓
Family name</td><td>Hu</td><td>名
First Name</td><td colspan="3">Yitong</td></tr>
<tr><td>BUPT 学号
BUPT number</td><td>2020213350</td><td>QM 学号
QM number</td><td>200980434</td><td>班级
Class</td><td>2020215111</td></tr>
<tr><td>论文题目
Project Title</td><td colspan="5">Design and development of a human-agent collaboration model for situation awareness in cockpit</td></tr>
<tr><td colspan="6">Please record supervision log using the format below:

Date: dd-mm-yyyy
Supervision type: face-to-face meeting/online meeting/email/other (please specify)
Summary:</td></tr>
<tr><td colspan="6">Date: 09-11-2023
Supervision type: online meeting
Summary: discussed the project generally

Date: 16-11-2023
Supervision type: online meeting
Summary: discussed the project specification

Date: 20-11-2023
Supervision type: email
Summary: received written feedback on the draft specification

Date: 23-11-2023
Supervision type: online meeting
Summary: discussed the specific work plan

Date: 27-11-2023
Supervision type: phone
Summary: discussed the specific tasks setting

Date: 30-11-2023
Supervision type: phone
Summary: revised the tasks defined in project specification

Date: 07-12-2023
Supervision type: online meeting
Summary: discussed the requirements for smart cockpit through literature survey

Date: 21-12-2023
Supervision type: online meeting
Summary: discussed the demo for demonstrating the performance of the interactive machine-learning model</td></tr>
</table>

Date: 11-01-2024
Supervision type: online meeting
Summary: received written feedback on the draft early-term report

Date: 25-01-2024
Supervision type: online meeting
Summary: introduced the system architecture and discussed the improvement

Date: 08-02-2024
Supervision type: online meeting
Summary: discussed the simulator for the experiments

Date: 22-02-2024
Supervision type: online meeting
Summary: discussed the final report structure and the draft mid-term report

Date: 07-03-2024
Supervision type: online meeting
Summary: reviewed the mid-term report and provided feedback for revisions

Date: 21-03-2024
Supervision type: online meeting
Summary: discussed the integration of human-agent collaboration models into the simulator

Date: 04-04-2024
Supervision type: online meeting
Summary: tested the first prototype of the system in the simulator and discussed initial results

Date: 18-04-2024
Supervision type: online meeting
Summary: mock viva

Additional Appendices

SA module’s system prompt:

Input: 1. First-person perspective image with driver’s gaze heat map 2. Human driver control data: Contains the driver’s keyboard input, JSON format data 3. External environment data: includes vehicle information within a radius of 50 meters (relative coordinates, speed, acceleration and collision box, including ego vehicles), JSON format data 4. Historical data: summaries of previous conversations

Task: Combines image analysis and control input to generate a structured report detailing: **A. Driver’s gaze point analysis**: - Recognized entities and their descriptions (e.g., color, type, distance, apparent motion) **B. Control data interpretation**: - Driver’s current driving behavior - Driver’s underlying driving intention **C. external env data interpretation**: - Movement behavior of external vehicles - Potential movement of external vehicles **D. Comprehensive analysis**: - Combined with the above analysis, describe the driver’s overall state, attention and driving intentions, and how external entities may affect the driver

Note: - Consider the relationship between the driver’s gaze point and the recognized entity - Analyze the consistency of driver control inputs with driving environment and gaze points - Evaluate how a driver’s physical state and attention level affect driving performance and safety - Utilize historical data comparisons to identify changes or consistency in behavioral patterns

Example output:

A. Driver’s gaze point analysis: - Entity: red car, about 30 meters away. - Entity: traffic light, showing red light.

B. Control data interpretation: - Current driving behavior: slow down slightly and prepare to stop. - Potential driving intent: Possibly preparing to stop at a traffic light. C. external env data interpretation: - The red car (eid=v-12345) has stopped and is waiting to pass at the intersection D. Comprehensive analysis: - The driver is currently focused on the road conditions and surroundings, preparing to stop at a red light. Control inputs are consistent with the environment ahead, showing good driving response and intent. Based on historical data, drivers generally demonstrate good concentration and driving skills in similar situations.

Risk and environmental impact assessment

In order to ensure the successful completion of the project, it is crucial to assess potential risks and environmental impacts. This assessment involves considering factors that could prevent project success, cause harm to people or animals, damage the environment, or lead to financial losses. Each risk is evaluated based on its likelihood of occurrence (L) and the seriousness of its consequences (C). The risk level (R) for each event is calculated as follows:

$$R = L \times C$$

This formula gives a numeric estimate of the risk level, guiding the necessary contingency planning. The following tables provide the scores for likelihood and consequence levels, which help in assessing the overall risk.

5.4 Likelihood Scores

Table 2 outlines the scores for different levels of likelihood:

Level L	Description	Meaning
0	Impossible	Cannot happen
1	Rare	May happen in exceptional circumstances
2	Unlikely	Could happen at some time
3	Moderate	Should happen at some time
4	Likely	Will happen often
5	Certain	Expected to happen

Table 1: Scores for level of likelihood

5.5 Consequence Scores

Table 3 provides the scores for the level of consequences:

Level C	Description	Meaning
0	Negligible	No noticeable effect on the project.
1	Minor	Undesirable but can be handled without major issues.
2	Serious	Might cause slight disruption.
3	Very Serious	Will cause significant disruption.
4	Major	Project likely irrecoverable in parts.
5	Catastrophic	Project completion impossible.

Table 2: Scores for level of consequence

5.6 Assessment of Risk Levels

Combining the likelihood and consequence scores from Tables 2 and 3, we assess the overall risk levels as shown in Table 4:

CL	**0**	**1**	**2**	**3**	**4**	**5**
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	6	8	10
3	0	3	6	9	12	15
4	0	4	8	12	16	20
5	0	5	10	15	20	25

Table 3: Assessed level of risk combining consequence and likelihood

5.7 Action Required Based on Risk Rating

Based on the risk scores, actions are recommended as outlined in Table 5:

Score	**Rating**	**Action**
0	No Risk	No action required.
1 - 3	Low Risk	Take action if easy to implement.
4 - 6	Moderate Risk	Take action if cost effective.
8 - 12	Significant Risk	Take action urgently.
15 - 25	High Risk	Requires immediate action.

Table 4: Ratings of risk and urgency of required action

5.8 Conclusion

This assessment provides a structured approach to identifying and managing risks in the project. By evaluating the likelihood and consequences of potential risks, we can prioritize actions and ensure that appropriate measures are in place to mitigate these risks, thereby safeguarding the project’s successful completion and minimizing environmental impact.