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: 11-04-2024

Table of Contents

Abstract

-> TBD

Keywords

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

摘要

-> TBD

关键词

人机协作、上下文感知、态势感知、隐式交互

Chapter 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 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.

This project introduces HarmonyCockpit (HCockpit), an agent architecture designed to incorporate cutting-edge large multimodal 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-awareness to facilitate HAC. Subsequently, HCockpit translates tasks into actions adapting to 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), the HCockpit serves as the orchestrating force between human and agent.

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 (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).

2.3 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.

Chapter 3: Design and Implementation

3.1 Preliminaries

Environment, Entity and Agent The environment is composed of all entities in it, and there are interactions between entities. This article defines the global environment 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 includes devices defined by the cockpit model (will be discussed later) , the autonomous driving system (AI driver), the human driver, and the AI copilot. However, entities have different levels of intelligence. Some entities can only respond mechanically to stimuli from the environment, while others can respond to the environment through Observation forms a certain understanding, and then autonomously plans certain tasks and performs a series of actions to pursue certain goals. In order to distinguish the two, this article calls the latter agent. From the agent's perspective, other non-agent entities in its environment are perceptible or controllable objects. Of course, the agent's ability to perceive or control other objects depends on itself. In addition, there may be multiple agents in an environment. Agents with higher intelligence levels are usually composed of complex systems, so it may be difficult for them to directly control each other. Instead, they need to convey their intentions to each other through communication to achieve a certain goal. Within the global environment, there are three types of agents: 1. AI driver, 2. human driver, and 3. AI copilot (an agent implemented using HCockpit). The first two are considered independent homomorphic agents, meaning they have the same observation space (the external environment). Differing from them, the AI copilot has global observation, which means it can perceive all entities in both the internal and external environments.

Action Action refers to the operation performed by the agent on entities in the environment, aiming to achieve its goals or respond to changes in the environment. In this article, since AI driver and human driver are homomorphic agents, they share the same action space, which is the control to the motion controller of the vehicle. The AI copilot, however, has a different action space, which is to control the devices (including the motion controller) in the cockpit. It is worth noting that in real driving situations, human drivers may not only control the motion controller, but also improve the driving experience by controlling cockpit equipment such as instrument panels, displays, and air conditioners. . The HCockpit architecture was designed with such situations in mind (see **Cockpit Model** for details), but in order to simplify the problem and quickly verify the usability of HCockpit, this article limits the human driver's actions to the control of the motion controller.

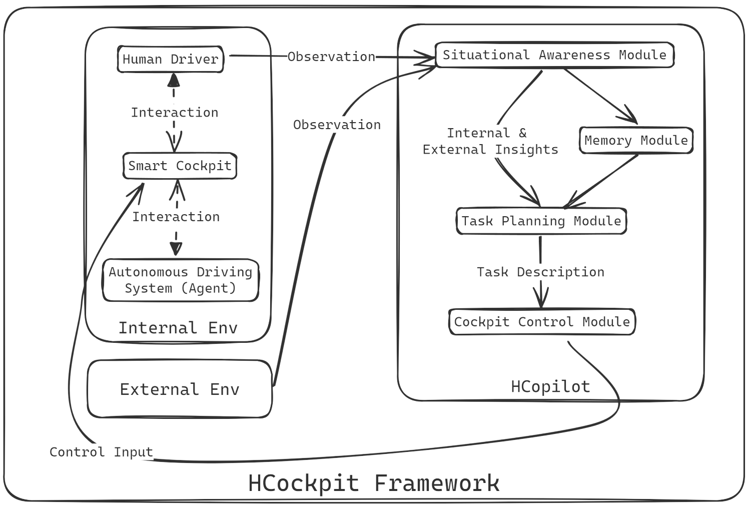

Figure 1: HCockpit Architecture

HAC **H**uman-**A**gent **C**ollaboration refers to the process in which the two work together to achieve a certain goal in a shared environment. In this article, HAC is the ultimate goal of AI copilot, which refers to the AI driver and the human driver working together to achieve a safe, comfortable and efficient driving experience in a shared driving environment. In order to achieve HAC, AI copilot needs to decompose this complex goal into a series of simple sub-tasks (HAC tasks), such as: 1. Using the AI driver's environmental awareness capabilities to enhance the driver's situational awareness, 2. Combining human drivers with The status provides driving style suggestions for the AI driver. Since the AI copilot has no physical entity, cockpit devices will serve as tools (embodiments) for it to perform HAC tasks, which requires the AI copilot to have the ability to control cockpit devices.

Cockpit Model This article introduces the cockpit model as a simplified model of the cockpit environment. The Cockpit model defines the functions of all devices in the cockpit and the control permissions of different roles (agents) on these devices. Technically, the cockpit model provides a unified application programming interface (API). In this way, AI copilot can fine-tune the functionality of cockpit devices to complete HAC tasks. The cockpit model is designed to be simple and easy to understand, so it can be used as a testbed to evaluate the HCockpit architecture. It is worth noting that the cockpit model does not involve the control details of specific devices, but designs a simple Device control interface available for evaluation and demonstration.

Figure[1] shows the architecture of HCockpit.

3.2 HCockpit Architecture

Situational Awareness Module 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 (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.

Context Awareness Module -> TBD

Planning Module 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.

Control Module 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.

3.3 HCopilot Implementation

-> TBD

To assess the efficacy of the HCockpit and garner deeper understandings, this project developed HarmonyCopilot (HCopilot) agent as an instance of the HCockpit architecture by leveraging state-of-the-art large multimodal models and incorporating advanced smart cockpit models, which will be discussed in detail below.

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.

3.3.1 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:

3.3.2 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:

3.3.3 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:

Chapter 4: Results and Discussion

Most projects will have results, especially for a research project. But again you should talk to your supervisor about it.

Chapter 5: Conclusion and Further Work

The conclusion is an important part of the report, as it states what you have done for the project.

5.1 Conclusion

It concludes the findings of your research or the outcome of implementing a system. A good conclusion will NOT repeat what you have done, but set out the achievements very crisply.

5.2 Reflection

This chapter should include a reflection statement for your project. You should critically reflect upon the technical skills developed, new knowledge gained, and lessons learnt on the project journey. You are encouraged to include reflections on broader ethical, social, legal, and environmental issues, allied with good professional practice and behaviour you have adopted in conducting your project.

5.3 Further work

Further work can be the next step of your research, or some functionality that can be added to the implementation to make it more practical.

NOTE: The maximum length of the report up to here is 50 pages.

Here is a cite test in LaTeX: (1)

References

[1] Ling CX, Yang Q. Crafting Your Research Future: A Guide to Successful Master's and Ph.D. Degrees in Science & Engineering. Synthesis Lectures on Engineering. Cham: Springer International Publishing; 2012. Available from: https://link.springer.com/10.1007/978-3-031-79351-6.

Acknowledgment

Give your acknowledgment to people who helped you during the project here. Maximum length of this section is 1 page. You may thank your supervisor but DO NOT MENTION YOUR SUPERVISOR'S NAME HERE.

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.:

QM No.:

Full Name (Pin Yin):

Full Name (Chinese):

Signature:

Date:

Project specification

Include your project specification, part 1 and part 2 here. It must be the final version submitted to QMPlus.

Early-term progress report

Include your project early-term progress report here. It must be the final version submitted to QMPlus.

Mid-term progress report

Include your project mid-term progress report here. It must be the final version submitted to QMPlus.

Supervision log

Include your project supervision log here.

Additional Appendices (as needed)

Information that you think may be helpful or relevant for the reader but that is not directly relevant to the story of your project. Things that might be suitable as an appendix to a report are:

- Large tables of numerical results that have been displayed graphically in the main body of the report.
- Important parts of datasheets for specific devices you have used in your project if you think that they are important enough that the reader should have access to them without finding them off the web themselves.
- Mathematical proofs and results that are important to show but not important to the flow of the story in the report.

NOTE: Full code listings must NOT be included as an appendix, but extracts of code may be included in the body of the report to illustrate particular points. Code should be submitted as supporting documents to QM+.

Risk and environmental impact assessment

Please refer to the project handbook section 3.6.12