VISUAL ATTENTION-BASED PILOTS' STATUS MONITORING AND PERFORMANCE ENHANCEMENT TOWARDS HUMAN-IN-THE-LOOP AUTOMATION

LYU MENGTAO

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

Department of Aeronautical and Aviation Engineering

Visual Attention-based Pilots' Status Monitoring and Performance Enhancement towards Human-In-The-Loop Automation

LYU Mengtao

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Nov 2024

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_________________________ (Signed)

LYU Mengtao (Name of student)

Dedication

Finding answers can be an exhilarating journey, filled with unexpected twists and frustrations. This thesis is dedicated to the researchers who have experienced, are experiencing, and will experience the journey.

Abstract

Automation systems have increasingly been implemented in the aircraft's cockpits to facilitate human pilots and reduce their workload. While automation systems have significantly improved operational efficiency and reduced human errors, poorly designed human-computer interaction (HCI) in cockpits has also introduced new challenges in the highly automated control loops, such as the impaired cognitive status induced by Out-Of-The-Loop (OOTL) phenomenon and the overload of the In-The-Loop (ITL) pilots caused by the multiple information resources in the cockpits. The existing strategies mostly focus on solely identifying the pilot's OOTL status with machine learning methods or providing assistance with predefined protocols based on the pilot's workload levels. However, these methods lack explainability and fail to offer proactive automation support specifically tailored to pilots' cognitive states and task demands. Therefore, closed-loop support is required to provide human-centric assistance to both the OOTL and ITL pilots with user-friendly interactions.

This thesis presents a systematic approach leveraging eye-tracking technology and artificial intelligence (AI) to enhance closed-loop support for pilots. This work addresses three key research problems, with contributions as follows. First, a novel Flashlight model integrates attention distribution and attention resource

metrics, providing a comprehensive framework to analyze pilots' visual attention and predict operational performance. Second, the Visual Attention LTL_f for Identifying OOTL (VALIO) framework employs linear temporal logic and graph neural networks to identify OOTL status with enhanced explainability, offering human-readable insights into pilots' behaviours. Third, an innovative Large Language Model (LLM)-based method tokenizes eye-tracking data into Visual Attention Matrices (VAMs) to detect and support ITL troubleshooting behaviours, enabling context-aware and resource-efficient human-computer interactions.

Several case studies were conducted at the Human Factors Lab in the Department of Aeronautical and Aviation Engineering (AAE) to verify the efficacy of the proposed methods. The eye-tracking measurements developed based on the Flashlight model improved the prediction accuracy of pilots' operation performance. The VALIO framework achieved a stable identification accuracy across different time windows, with F1 scores around 0.8. And the explainability is significantly increased by the generated human-readable formulas. The integration of eye-tracking techniques and LLM achieved a Micro-average F1 score of 0.852 for identifying where the pilot is troubleshooting, with proactive and user-friendly interactions.

In conclusion, this thesis contributes to aviation safety by developing innovative methods for monitoring, predicting, and supporting pilot performance in both OOTL and ITL scenarios, advancing the human-in-the-loop HCI in modern cockpits. These developments lay the groundwork for safer and more efficient aviation operations.

Publications arising from the thesis

Journal articles (in reverse chronological order)

[1] **Lyu, M.**, Li, F., Lee, C.-H., Chen, C.-H. "VALIO: Visual attention-based linear temporal logic method for explainable out-of-the-loop identification". In: *Knowledge-Based Systems* 299 (Sept. 2024), p. 112086. ISSN: 09507051. DOI: 10.1016/j.knosys.2024.112086. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950705124007202 (visited on 06/22/2024).

[2] **Lyu, M.**, Li, F., Qu, X., Li, Q. "Flashlight model: Integrating attention distribution and attention resources for pilots' visual behaviour analysis and performance prediction". In: *International Journal of Industrial Ergonomics* 103 (Sept. 2024), p. 103630. ISSN: 01698141. DOI: 10.1016/j.ergon.2024.103630. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169814124000866 (visited on 09/30/2024).

[3] **Lyu, M.**, Fan, L., Gangyan, X., Su, H. "Leveraging eye-tracking technologies to promote aviation safety- A review of key aspects, challenges, and future perspectives". In: *Safety Science* 168 (2023), p. 106295. ISSN: 0925-7535. DOI: https://doi.org/10.1016/j.ssci.2023.106295.

URL: https://www.sciencedirect.com/science/article/pii/S0925753523002370.

Manuscript under review

[1] **Lyu, M.**, Li, F. “Do you need help? Identifying and responding to pilots’ troubleshooting through Eye-Tracking and Large Language Model”. In: *International Journal of Human-Computer Studies (under review)* (Oct. 2024).

Conference paper (in reverse chronological order)

[1] **Lyu, M.**, Li, F. “Patterning Risk: An Innovative Task Design Method for Simulating Incidents in Transportation Studies”. In: 15th International Conference on Applied Human Factors and Ergonomics (AHFE 2024). 2024. DOI: 10.54941/ahfe1005198. URL: https://openaccess.cms-conferences.org/publications/book/978-1-964867-24-3/article/978-1-964867-24-3_9 (visited on 11/01/2024).

[2] **Lyu, M.**, Li, F., Lee, C.-H. “The Effects of Adaptive Automation on Pilots’ Flight Control Performance and Visual Attention Distribution”. In: *Advances in Transdisciplinary Engineering*. Ed. by Pisut Koomsap, Adam Cooper, and Josip Stjepandić. IOS Press, Nov. 7, 2023. ISBN: 978-1-64368-440-6 978-1-64368-441-3. DOI: 10.3233/ATDE230600. URL: https://ebooks.iospress.nl/doi/10.3233/ATDE230600 (visited on 11/24/2023).

Acknowledgements

I would like to attribute the completion of this thesis to many people who have inspired me, encouraged me, and supported me.

First and foremost, I extend my heartfelt gratitude to my chief supervisor, Dr. Li Fan, for her patient guidance, insightful suggestions, inspiring encouragement, and unwavering support throughout the completion of my work. Her conscientious research attitude, modest academic spirit, and optimistic philosophy towards life have profoundly influenced me, not only as a supervisor but also as a friend and life mentor. I would also like to send my sincere gratitude to my co-supervisor, Dr Xu Gangyan, whose invaluable guidance and encouragement have been instrumental in my research and thesis. Their dedication has made this accomplishment possible.

I also wish to express my sincere gratitude to my master's supervisor, Professor Chen Chun-Hsien, and my advisors, Dr. Li Xinyu, Dr. Wang Zuoxu, Dr. Zheng Pai, Dr. Kam K.H. Ng, and Dr. Li Qinbiao. Their direct instruction in improving my research methods and writing skills, along with their academic guidance, has been invaluable. They have illuminated my career path in academia since my master's studies, and their mentorship will always be cherished in my future academic endeavors.

My gratitude extends to my cherished teammates in the “SHINE group” and my friends who have accompanied me throughout my PhD studies. The fruitful discussions, helpful suggestions, generous assistance, and joyful laughter they provided have been a great source of encouragement and spiritual support in both my research and everyday life.

Last but not least, I am deeply grateful to my parents for their limitless love and unwavering support. They have been, and will always be, the driving force behind my journey to becoming a better person.

Contents

Abstract iii

Publications arising from the thesis v

Acknowledgements vii

List of Figures xii

List of Tables xiii

Nomenclature xiv

1 Introduction **1**

1.1 Background 1

1.2 Objectives 6

1.3 Scope 7

1.4 Significance 9

1.5 Structure of the thesis 9

2 Literature Review **12**

2.1 Human-computer interaction (HCI) in aviation 13

2.1.1 Out-Of-The-Loop (OOTL) and In-The-Loop (ITL) statuses 13

2.1.2 Identifying OOTL and ITL 14

2.2 Eye-Tracking techniques in aviation studies 15

2.2.1 Development 15
2.2.2 Remaining challenges 17
2.2.2.1 Enhancing the understanding of visual attention 17
2.2.2.2 Enhancing explainability of status identification 19
2.2.2.3 Enhancing the interaction for ITL support . . . 20
2.3 Linear Temporal Logic (LTL) 21
2.4 Processing psychological data with Large Language Model (LLM) 21

3 Study 1: A Visual attention analysis model 24

3.1 Background 25
3.2 Flashlight model 26
3.2.1 Visual attention metrics in human performance 26
3.2.2 Development of Flashlight model 28
3.2.3 Eye-tracking metrics based on Flashlight model 29
3.3 Case study 31
3.3.1 Participants and apparatus 32
3.3.2 Experiment procedure and task design 33
3.3.3 Flight control performance 34
3.3.4 Eye-tracking measurements 37
3.3.5 Data analysis and comparison 39
3.4 Results 42
3.4.1 Performance of different metrics 42
3.4.2 Discussions 44
3.4.2.1 Explanations of the results 45
3.4.2.2 Implications 46

3.4.2.3 Summary 47
3.5 Concluding remarks 47

4 Study 2: A explainable OOTL identification framework 50

4.1 Background 51
4.1.1 Identifying the OOTL status 51
4.1.2 Linear Temporal Logic for behaviour classification 53
4.1.3 Parse visual attention through eye-tracking 54
4.1.4 Challenges of applying LTL methods 55
4.2 Visual Attention LTL_f for Identifying OOTL (VALIO) 56
4.2.1 Visual Attention Traces (VATs) 57
4.2.2 Linear Temporal Logic on finite traces (LTL_f) 60
4.2.3 OOTL Identification 62
4.3 Case study 63
4.3.1 Experiment design and data collection 63
4.3.2 Learn LTL_f formulae with Visual Attention Traces 65
4.3.2.1 Areas of Interests in the cockpit 66
4.3.2.2 Divide dwell times into different levels 67
4.3.2.3 Encoding VATs 68
4.3.3 Results and comparisons 69
4.3.3.1 Identified LTL_f formulas 70
4.3.3.2 Sensitivity of time windows and division methods 72
4.3.3.3 Comparison with other methods 73
4.3.4 Discussion 74
4.3.4.1 VAT lengths and dwell time divisions 76

4.3.4.2 VALIO and other classification methods 77

4.4 Concluding remarks 78

5 Study 3: A context-aware ITL support approach 80

5.1 Background 81

5.2 Visual Attention Matrix (VAM) 83

5.2.1 Compiliing eye-tracking data into VAM 83

5.2.2 Detecting pilots' information needs with LLM and VAM . 85

5.3 Case study 89

5.3.1 Experiment design and data collection 89

5.3.1.1 Apparatus and participants 89

5.3.1.2 Experiment design 90

5.3.2 Data Processing 93

5.3.3 Results 97

5.4 Discussion 100

5.4.1 Implications derived from the results 101

5.4.2 Limitations 102

5.5 Concluding remarks 103

6 Conclusion 105

6.1 Contribution 105

6.2 Limitations and future research 107

References 110

Appendices 151

List of Figures

1.1 Automation support based on human pilot's status 3
1.2 Organization of the thesis 11
3.1 Flashlight model for visual attention 27
3.2 Combined AOI-gaze metrics 31
3.3 Apparatus and experiment settings 33
3.4 Experiment flight process 35
3.5 Pitch and row angle during descent 37
3.6 Areas of Interest (AOIs) in the flight simulator 38
3.7 Data analysis procedure 40
3.8 Prediction accuracy with selected features 43
3.9 Variation of prediction results from actual performance 44
4.1 Framework of Visual Attention LTL_f for Identifying OOTL (VALIO) 57
4.2 Composition of LTL_f traces and VAT 60
4.3 Experiment apparatus 64
4.4 Flight tasks 66
4.5 Screenshot of Tobii Pro Lab 67
4.6 Division of dwell times using different methods 69
4.7 Temporal-spatial information in the VATs 70

4.8 Distribution and performance of LTL_f formulas 71
4.9 VALIO performance using different dwell time division methods . 73
4.10 Performance of VALIO compared with other methods 74
5.1 Construct Visual Attention Matrix (VAM) from eye-tracking data . 84
5.2 Detecting pilots' information needs with GPT-4 and VAM 88
5.3 Experiment apparatus 90
5.4 Flight route and tasks 92
5.5 9x9 VAMs for the three tasks (Participant 3, RJBB to RJAA) . . . 94
5.6 Classification confusion matrix of different input 98
5.7 Macro and Micro evaluation metrics using different inputs 99

List of Tables

3.1 Top 5% measurements in each input group 42
4.1 Machine learning models and the major hyper-parameters 75
5.1 Expert scoring result 100

Nomenclature

AI	Artificial Intelligence
AOI	Area of Interest
ASRS	Aviation Safety Reporting Systems
ATCO	Air Traffic Control Officer
ECAM	Electronic Centralized Aircraft Monitor
EEG	Electroencephalogram
FCU	Flight Control Unit
FMS	Flight Management Systems
FN	False Negative
fNIRS	functional Near-Infrared Spectroscopy
FO	First Officer
FP	False Positive
GBDT	Gradient Boosting Decision Tree
GNN	Graph Neural Network
HCI	Human-Computer Interaction
IDA	Information-Decision-Action (cognitive model)
ITL	In-The-Loop
LDA	Linear Discriminant Analysis

LDG	Landing Gear
LightGBM	Light Gradient Boosting Machine
LLM	Large Language Model
LOA	Level of Automation
LTL	Linear Temporal Logic
LTL_f	Linear Temporal Logic on finite traces
MaxSAT	Maximum Satisfiability problem
MCDU	Multifunction Control Display Unit
MCI	Mild Cognitive Impairment
MFD	Multi-Function Display
MLP	Multilayer Perceptron
NASA	National Aeronautics and Space Administration
NASA-TLX	NASA Task Load Index
ND	Navigation Display
OOTL	Out-Of-The-Loop
PFD	Primary Flight Display
RF	Random Forest
RVR	Runway Visual Range
SA	Situation Awareness
SAGAT	Situation Awareness Global Assessment Technique
SPO	Single Pilot Operation
TCAS	Traffic Collision Avoidance System
TN	True Negative
TP	True Positive

UAV	Unmanned aerial vehicles
VALIO	Visual Attention LTL_f for Identifying OOTL
VAM	Visual Attention Matrix
VAT	Visual Attention Trace
VR	Virtual Reality
XGBoost	eXtreme Gradient Boosting
8-a	8x8 VAMs without P-Value VAM
8-Sum	fixation durations on the 8 AOIs
8-p	8x8 VAMs with P-Value VAM
9-a	9x9 VAMs without P-Value VAM
9-Sum	fixation durations on the 8 AOIs, and sum total
9-p	9x9 VAMs with P-Value VAM

Chapter 1

Introduction

This introductory chapter comprises five sections. Section 1.1 provides the research background on human-in-the-loop automation in aviation, emphasizing the advantages of utilizing visual attention in aviation studies and outlining the motivation for this work. Section 1.2 presents the research objectives, derived from the background. Section 1.3 defines the research scope, followed by Section 1.4, which discusses the significance of the work. Finally, Section 1.5 outlines the organization of the thesis.

1.1 Background

Human errors are well recognized as a major cause of aviation accidents, contributing to over 70% of such incidents due to factors like fatigue, mind wandering, and reduced situational awareness [1–3]. Automation systems have increasingly been implemented to enhance efficiency and reduce human workload [4]. The introduction of advanced autopilot systems and flight management functionalities has significantly reduced human errors by automating tasks traditionally carried out by pilots [5]. Autopilots and Flight Management Systems (FMS) automate multiple flight operations, such as control, navigation, information display, and fuel management. While automation enhances efficiency and reduces workload, optimizing human-computer interaction (HCI) remains crucial in highly automated settings. Poorly designed HCI, particularly in high levels of automation (LOA), can lead to severe failures when human operators struggle to manage or intervene effectively [6, 7].

Human-automation interaction has been studied extensively, leading to the categorization of multiple levels of human involvement in control tasks: *in-the-loop*, *on-the-loop*, and *out-of-the-loop* [8, 9]. ”*In-the-loop*” indicates that the human operator actively participates in real-time control and decision-making, while ”*on-the-loop*” refers to a supervisory role in which humans monitor and are prepared to intervene if required. In this study focused on pilots, both statuses are considered ”*in-the-loop*” (ITL), as both require active monitoring, which is more common in modern commercial aviation compared to manual control. On the other hand, ”*out-of-the-loop*” (OOTL) refers to situations where the human operator is disengaged from monitoring and decision-making, often leading to decreased situational awareness and performance deterioration during critical interventions [10, 11]. Both NASA data and surveys among German aviators indicate that the OOTL phenomenon significantly contributes to human errors in aviation [12, 13]. Meanwhile, ITL pilots may also face overload due to high workloads from concurrent tasks, posing significant safety risks [14, 15]. Hence, adaptive automation support tailored to the human pilot’s status is required.

As illustrated in Figure 1.1, automation support demands vary based on the pilot's status. To deliver closed-loop support, automation systems must provide timely warnings to re-engage OOTL pilots in the control loop or autonomously manage non-critical tasks. For ITL pilots, the automation system must assist with troubleshooting by providing critical information and supporting decision-making. This study aims to develop a systematic method for delivering closed-loop support tailored to pilots' status and their specific needs for support.

Figure 1.1: Automation support based on human pilot's status

To monitor pilot status and understand their needs for support, early studies employed questionnaires and scales, such as the NASA Task Load Index (NASA-TLX) [16, 17] and the Situation Awareness Global Assessment Technique (SAGAT) [18, 19]. Despite being effective, these methods are limited by response bias and can interfere with operations. To provide continuous and objective monitoring, recent studies have shifted to physiological data collection using methods such as Electroencephalography (EEG) [20, 21], functional Near-Infrared Spectroscopy (fNIRS) [22, 23], heart rate monitoring [24], and eye-tracking [25, 26]. Eye-tracking, in particular, has shown significant promise for practical application. Unlike other techniques, such as EEG or fNIRS, eye-tracking provides actionable and explainable insights, enabling targeted interventions to enhance pilots' focus on key information sources. Moreover, its non-invasive nature makes it advantageous over other biometric approaches. A recent review highlighted eye-tracking as the most effective method for assessing mental states in supervisory control tasks [27]. Therefore, this work employs eye-tracking technologies to capture pilots' visual attention and analyze their status.

Despite the widespread adoption of eye-tracking technologies in aviation studies, challenges remain in developing a visual attention-based framework for pilot status monitoring and closed-loop automation support. These challenges primarily concern understanding visual attention, ensuring the explainability of estimation results, and optimizing human-computer interaction, as outlined below:

- **Understanding pilots' visual attention**: Existing research uses eye-tracking to predict human performance from two perspectives: metrics focused on specific Areas of Interest (AOIs) to assess attention distribution, and general metrics indicating available cognitive resources for processing information. Metrics such as total fixation duration, fixation count on specific AOIs, and transitions between AOIs reflect the attention allocated to particular stimuli [28–30]. Meanwhile, Attention resource metrics like pupil diameter and saccade velocity are used to evaluate the cognitive status and information processing efficiency [31, 32]. Most studies analyze these perspectives independently, but a more integrated approach is required for a comprehensive understanding of pilots' visual attention.

- **Explainability of estimation results**: Many existing studies rely on machine learning models to analyze pilots' eye-tracking data and estimate their status [33, 34]. The "black-box" nature of these methods raises concerns in risk-sensitive domains like aviation, where opaque decision-making can lead to severe consequences [35, 36]. Incorrectly identifying OOTL status can result in unwarranted alarms, leading to alarm fatigue and impairing pilot performance [37]. Enhancing the explainability of these estimations is critical, enabling more effective, human-centred alerts that help pilots recognize and adjust their behaviour to mitigate the risks of OOTL situations.

- **Interaction approach**: Most existing approaches directly adjust automation functions based on pilots' status and predefined protocols, often disregarding the pilot's actual needs. This can lead to inadequate support or unnecessary interventions, causing alarm fatigue [38]. Closed-loop automation support should consider the pilot's intentions and deliver user-friendly interactions. Communication between the system and the pilot is essential to confirm support requirements and improve effectiveness.

This thesis presents a formal study to address these three challenges. The work adopts Artificial Intelligence (AI) techniques, such as machine learning, Graph Neural Network (GNN), Linear Temporal Logic (LTL), and Large Language Model (LLM), etc., to develop a systematic method to provide close-loop support to the pilots based on visual attention.

1.2 Objectives

The primary objective of this thesis is to develop a systematic approach for closed-loop support tailored to pilots, comprehensively considering their cognitive status and task demands based on visual attention. Based on this overall goal and the three major challenges, this goal is separated into three sub-objectives:

- *Model the pilot's visual attention and cognitive status*

  Modelling the pilot's visual attention with collected eye-tracking data is the first step in analysing the pilot's cognitive status and identifying the critical measurements to predict the pilot's performance [27]. Generally, the eye-tracking measurements can reflect human's visual attention from two perspectives: metrics focused on specific Areas-Of-Interest (AOIs) to reflect attention distribution, and general metrics implying the amount of available attention resources for information processing tasks.[28]. To enhance the granularity of eye-tracking data analysis, the primary goal of this work is to model the pilot's visual attention by integrating both attention distribution and attention resource perspectives.

- *Identify pilots OOTL status with explainability*

  Explainable identification of Out-of-the-Loop (OOTL) status is vital for effective closed-loop automation support [8, 10]. Interpretable results enable the system to generate specific warnings or apply appropriate automation assistance. To enhance the explainability of OOTL identification results, this work adopts the Linear Temporal Logic (LTL) methods [39] to analyse the pilot's eye-tracking data. To adapt the eye-tracking data with the LTL methods, we proposed a method to encode pilots' visual attention distribution data into formalized traces and developed a framework, named *Visual Attention LTL_f for Identifying OOTL (VALIO)*. This framework segments the cockpit view into several distinct AOIs based on functionality and spatial organization, assessing pilots' gaze direction and attention distribution with both temporal and spatial considerations.

- *Capture ITL pilot's task demands and proactively respond with support*

  Another aspect of close-loop automation support for human pilots is to detect the demand for support from the ITL pilot and proactively respond to it. ITL pilots often experience high workloads, especially in scenarios requiring attention to multiple concurrent tasks [40, 41]. In unexpected situations, pilots must troubleshoot while simultaneously monitoring flight status, leading to a challenge due to the limited field of view (approximately 4 degrees) [28, 42]. Hence, this work proposes a method to detect and respond to the troubleshooting activities of ITL pilots using the Large Language Model (LLM).

1.3 Scope

Concerning the research objectives, this work focuses on understanding the pilot's visual activities, cognitive status, and task demands to establish a systematic approach for closed-loop automation support. The scope of the research is as follows:

- *A model to evaluate the pilot's cognitive status*

This model integrates the attention distribution and attention resources perspectives of the eye-tracking data to evaluate the pilot's cognitive status and predict performance. Therefore, eye-tracking measurements with higher granularity that combine these two perspectives will be introduced. Furthermore, the most impactful measurements that correlated to the pilot's performance will be identified.

- *A framework to identify pilots OOTL status with explainability*

  The framework needs to not only identify the pilots' OOTL status but also enable explicit explanations. To achieve this, the eye-tracking data in certain time windows will be segmented to compile LTL traces. The Graph Neural Network will be adopted to analyse the embedded temporal-spatial information and obtain LTL formulas. These obtained human-readable LTL formulas facilitate the explanations of the OOTL identification results and provide more insights into the visual behaviours in OOTL status.

- *An AI co-pilot to detect and respond to ITL pilot's troubleshooting*

  Leveraging LLMs, the AI co-pilot processes eye-tracking data for comprehensive and context-aware troubleshooting detection. A novel method to tokenize pilots' eye-tracking data for LLM processing will be introduced. Meanwhile, empirical data will be utilized to reduce the resource costs of LLM training. Beyond the identification of troubleshooting behaviour, the natural interaction capability and context-awareness of LLM will be leveraged to respond proactively to the pilot, facilitating user-friendly interactions to assist pilots.

1.4 Significance

By achieving these objectives, this research will make contributions to the following research areas:

- *Exploratory research on pilot's OOTL and ITL status in aviation*

  While the importance of HCI in highly automated cockpits has been gradually realized, most studies focus solely on either OOTL or ITL scenarios. This research is positioned to make a pioneering contribution by taking both statutes into consideration and specifying the different need for support corresponding to the statutes. This will lay a foundation for a more comprehensive understanding of pilots' needs during flights and advance HCI in cockpits to enhance aviation safety.

- *Development of a close-loop support scheme based on the pilot's visual attention*

  This scheme involves a thorough analysis of the pilot's different statuses, including the characteristics and the specific needs for support in each status. Novel methods for status discrimination and proactive support provision are developed. Hence, this research is expected to make academic contributions to the realization of close-loop automation support with comprehensive consideration of the pilot's status and demands.

1.5 Structure of the thesis

This thesis consists of six chapters, with the ouThe organization of this thesis is demonstrated in Figure 1.2.

Chapter 1 introduces the background, objectives, scope, significance and overall structure of this work.

Chapter 2 reviews the pilots statuses and eye-tracking studies in aviation, proposing the research challenges in achieving a closed-loop automation support with eye-tracking techniques. In response to the challenges, the Linear Temporal Logic (LTL) methods and Large Language models (LLMs) are proposed to be adopted. Therefore, the LTL studies and the LLM-based psychological data processing are also review in this chapter.

Chapter 3 proposes a novel Flashlight model that combines the attention resources and attention distribution to analyse the pilot's visual attention and cognitive status. A case study to verify the proposed model and proposed eye-tracking measurements is described. Meanwhile, the most impactful eye-tracking measurements that correlate to the aircraft control performance are identified.

Chapter 4 presents a Visual Attention LTL_f for Identifying OOTL (VALIO) framework. In this framework, the methods of linear temporal logic on finite traces (LTL_f) are utilised to identify the pilot's OOTL status with explainability. A novel method to encode the eye-tracking data into Visual Attention Traces (VATs) is introduced, and a GNN is leveraged to parse the VATs. A case study is conducted to test the effectiveness of the proposed method and find the optimal length of VATs.

Chapter 5 introduces an innovative approach to detect and respond to the ITL pilot's troubleshooting behaviours by utilizing the LLM. The eye-tracking data are tokenized into Visual Attention Matrices (VAMs) as LLM input. The context-aware ability of LLM facilitates capturing the complex troubleshooting status of pilots, which blends with normal monitoring behaviors. The use of empirical

VAMs contribute to effectively processing non-semantic eye-tracking data with LLM. Notably, the proposed method levearges the natural interaction capability of LLM to enable a proactive response by confirming the pilot's status and proposing to support. By conducting a verification case study, the performance of identifying troubleshooting behaviours is computed and the response messages are evaluated by a group of aviation experts.

Chapter 6 concludes the thesis with a summary of the research findings, discussions, and future research plans.

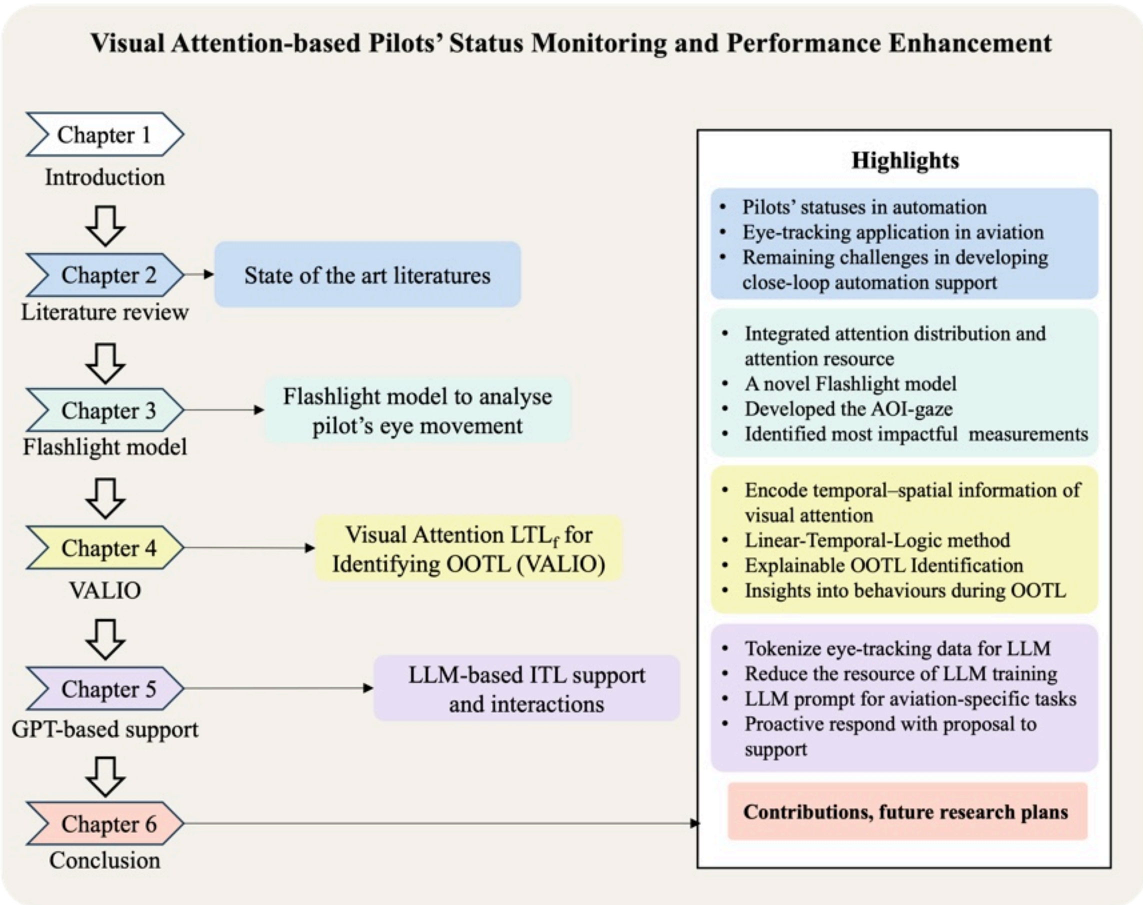

Figure 1.2: Organization of the thesis

Chapter 2

Literature Review

This chapter summarises the relevant studies on the techniques to be utilised in this study, and identifies their existing challenges in achieving a closed-loop automation support approach. First, Section 2.1 briefly reviews the human-computer interaction (HCI) and the pilot's cognitive status in aviaton. Then, the applications of eye-tracking techniques in aviation studies are presented in 2.2. Next, Section 2.3 illustrates the Linear Temporal Logic (LTL) methods to explore an explainable method for the pilot's status identification. Finally, section 2.4 introduces how the Large Language Models (LLMs) are used in processing psychological data.

2.1 Human-computer interaction (HCI) in aviation

2.1.1 Out-Of-The-Loop (OOTL) and In-The-Loop (ITL) statuses

Automation systems have become increasingly prevalent across various transportation sectors, especially within the aviation industry [4]. While automation improves efficiency and reduces workload, it also necessitates optimizing HCI in highly automated environments. Poorly designed HCI in high levels of automation (LOA) has been linked to severe failures when human operators are unable to effectively manage or intervene in automated systems [6, 7]. Therefore, researchers have extensively explored HCI in human-automation concurrent task environments and defined multiple levels of human engagement in the control loop: *in-the-loop*, *on-the-loop*, and *out-of-the-loop* [8, 9]. "*In-the-loop*" indicates that human operators are actively engaged in real-time control and decision-making tasks, while "*on-the-loop*" signifies a monitoring role, where humans supervise and are ready to intervene if necessary. In this study, we treat both statuses as "*in-the-loop*" (ITL) since both involve active monitoring, and modern commercial flights rarely require massive physical control. In contrast, "*out-of-the-loop*" (OOTL) indicates that the human operator is not engaged in monitoring or decision-making, leading to reduced situational awareness and potential performance degradation when intervention is required [10, 11].

The OOTL phenomenon carries substantial implications, from reduced decision and action performance, to critical incidents and accidents [12, 13]. Operators experiencing the OOTL phenomenon may find it challenging to promptly identify system malfunctions or failures [43]. Such delayed or inappropriate responses to system anomalies can further intensify the severity of potential accidents [44]. Research suggests that the OOTL phenomenon increases operators' workload when they are required to regain control after automation failure or disengagement, thereby placing additional strain on their attention resources [45, 46]. Meanwhile, the pilots' needs for support in ITL status cannot overlooked. The possible oversight of ITL pilots' needs for support may result in suboptimal performance or potential risks, particularly in scenarios requiring simultaneous attention to multiple tasks [40, 41]. For instance, pilots may find it challenging to visually troubleshoot an error while monitoring the flight status independently, given the limited field of view (approximately 4 degrees) [28, 42]. This increased workload of ITL pilots might also induce severe accidents [47].

2.1.2 Identifying OOTL and ITL

To assess pilots' cognitive status, early investigations introduced various questionnaires and scales, such as the NASA Task Load Index (NASA-TLX) [16, 17] and the Situation Awareness Global Assessment Technique (SAGAT) [18, 19], to assess these factors and their correlation with pilots' performance. However, while these methods are straightforward and effective, they are often constrained by respondent bias and can interfere with operations. In order to continuously and objectively monitor pilots' status and predict their performance, recent studies have frequently turned to collecting pilots' physiological data using techniques such as Electroencephalograms (EEG) [20, 21], functional Near-Infrared Spectroscopy (fNIRS) [22, 23], heart rate [24], and eye-tracking [25, 26].

Among these physiological measures, eye-tracking holds significant promise for practical implementation. A recent review has indicated that eye-tracking is the most widely adopted and effective method for assessing the mental state of human operators in supervisory process control tasks [27]. Investigating visual behaviours via eye-tracking technology offers a valuable window into pilots' attentional focus and cognitive status, which are key factors in detecting their status and needs for support [48, 49]. Unlike methods based on functional Near-Infrared Spectroscopy (fNIRS) and Electroencephalograms (EEG), which lack direct implications for behavioural adjustments, eye-tracking data provide actionable and explainable insights. Specifically, it enables the formulation of precise recommendations for pilots to enhance their engagement with critical information sources, thereby addressing the OOTL phenomenon (for example, advising increased focus on specific displays or instruments). And the pilots' visual attention distribution can reveal their current focus to provide insights into what support they may need. Moreover, eye-tracking technology facilitates non-invasive data collection, offering a significant advantage over other biometric approaches. Consequently, this study leverages eye-tracking to analyse pilots' visual attention, exploring the methods for identifying pilots' status and their needs for support.

2.2 Eye-Tracking techniques in aviation studies

2.2.1 Development

It has been over one hundred years since the eye-tracking approach was first adopted in aviation studies, and the eye-tracking technique has been widely studied in aviation as a proven approach now [50]. The pioneer researchers started to study the pilots' eye movements following body rotations via naked-eye observations [51] when aviation activities were almost only for military purposes at that time. Nowadays, commercial aviation activities have grown to become one of the most vital and complex industries that contribute over$2.7 trillion to the world's gross domestic product (GDP) [52] Meanwhile, eye-tracking technologies have experienced an emerging development with various techniques and algorithms to track the gaze position and direction automatically in real-time and have been widely applied to many areas, including HCI, gaming, and automatic safety research [53, 54]. Leveraging eye-tracking technologies to study eye movement has been widely adopted to improve the safety and human performance of aviation activities [31, 55].

The eye-tracking technologies capture signals from the movements and activities of the pupil, cornea, sclera, iris, retina, and other eye components by several different methods such as shape-based, appearance-based, feature-based, and hybrid methods [56, 57]. Now, both the hardware and the software of eye trackers have evolved owing to the comprehensive development of materials, sensors, artificial intelligence (AI), and other technologies [58]. In addition to the conventional eye-tracking techniques based on scleral search coil, electrooculography (EOG), infrared oculography (IOG), and video oculography(VOG), various rising advanced technologies, such as machine learning (ML), internet of things (IoT), cloud computing, has also been integrated to improve the eye-tracking results and explore eye-tracking applications [54]. Several studies and applications have confirmed that eye movements can act as effective and non-intrusive real-time performance indicators of visual attention and mental state in humans [59–

64].

2.2.2 Remaining challenges

As discussed in Section 1.1, there remain three challenges in monitoring pilot status and developing closed-loop automation support.

2.2.2.1 Enhancing the understanding of visual attention

Existing studies typically employ eye-tracking measurements to predict human performance from two perspectives: metrics focused on specific Areas-Of-Interest (AOIs) to reflect attention distribution, and general metrics implying the amount of available attention resources for information processing tasks.

- *Attention distribution metrics* commonly include total fixation duration, fixation count directed towards predefined AOIs, and transitions between these AOIs, indicating the level of interest and attention allocated to particular stimuli [28–30]. For example, research has shown that novice pilots tend to allocate more total fixation time to the airspeed indicator and heading indicator, whereas expert pilots distribute their fixation durations more evenly across all instruments, reducing the likelihood of overlooking critical information and thereby enhancing performance [65, 66]. Similarly, studies have identified that distractions can be detected by observing reduced viewing time on specific instruments during different flight phases (e.g., decreased attention towards the attitude indicator during cruising) [25].
- *Attention resource metrics*, such as pupil diameter, and saccade (peak) velocity, are commonly utilized to evaluate operators' cognitive status and

the efficiency of information processing at a holistic level [31, 32]. For instance, decreases in saccade velocity have been associated with increased fatigue in both simulated and real flight scenarios [67, 68]. Furthermore, pupil diameter has been demonstrated to be highly correlated with factors such as perceived workload and fatigue and has been employed in assessing pilots' performance in numerous studies [69–71].

While both metric types are widely adopted in aviation research, most existing studies apply them separately without further integration. For instance, Lutnyk et al. leveraged attention resource metrics (denoted as *AOI-independent metrics* in their work), including average fixation duration and saccade amplitude, to distinguish pilots' task statuses and normal flight conditions as they arouse different workload [72]. Meanwhile, their findings also adopted attention distribution metrics and confirmed that pilots devote more total fixation duration to AOIs that are closely associated with primary tasks during specific flight phases. However, more attention needs to be paid to the integration and comprehensive analysis of both attention distribution and attention resource two perspectives.

Human performance in complex and dynamic aviation human-machine collaboration environments is dependent on the information processing capability and how they distribute their attention simultaneously. Both impaired information processing and inefficient attention allocation across multiple tasks can hinder operators from achieving optimal performance [34, 73, 74]. For instance, a study in road traffic integrated the gaze trajectory (attention distribution) and pupil diameter (attention resources) to estimate the drivers' workload, and obtained superior performance than single modes [75]. Moreover, although pupil diameter is typically considered a general metric unrelated to specific AOIs, a study by Zhang et al. revealed significant variations in pupil diameter under different time and road complexity conditions in electronic centralized aircraft monitoring (ECAM), but not with other internal instruments [76]. This suggests that the granularity of eye-tracking data should be enhanced by combining both attention distribution and attention resource perspectives to enhance the understanding of pilots' visual attention. A Flashlight model with a consideration of both perspectives is proposed in Chapter 3.

2.2.2.2 Enhancing explainability of status identification

Generally, previous works have studied pilots' status from the perspective of data characteristics supported by machine learning methods. These machine-learning-based methods offer a robust capability for processing extensive and comprehensive data [26, 77]. However, there is an absence of explainability in the inference process of these deep learning or ensemble models with complex hierarchical structures since the rationale behind their decisions are hard to understand and interpret [78–80]. Therefore, these "black-box" methods has sparked significant concerns about their application in the high-risk aviation industry [35]. Meanwhile, the status identification results derived from purely digital numbers cannot provide direct guidance on how pilots should modify their behaviours to prevent or mitigate the negative effects. To address this gap, it is crucial to develop explainable methods for status identification based on eye-tracking data, and generate more instructive recommendations for pilots on how they should manage their visual behaviours.

This research innovatively integrates Linear Temporal Logic (LTL) and pilots' visual behaviours for the explainable identification of their OOTL status [39]. The LTL methods are reviewed in Section 2.3, and the LTL-based framework is presented in Chapter 4.

2.2.2.3 Enhancing the interaction for ITL support

Based on the analysis of eye-tracking and other physiological data, researchers develop methods to assess operators' status and provide support in different levels [61, 81]. A common strategy is adaptive automation, which dynamically adjusts the LOA based on operator's status [82, 83]. However, most existing literature has focused on mitigating risks associated with the OOTL phenomenon by adjusting automation levels, overlooking the more comprehensive interaction with ITL pilots. Considering the more dynamic needs of the pilots when they are actively engaged with ITL status, the oversight can result in suboptimal performance. For instance, redundant warning messages to the ITL pilots may disrupt the user and cause alarm fatigue [38]), while fully relying on the ITL pilots may lead to a lack of necessary assistance. This necessitates a more user-friendly interaction approach that is capable of understanding the ITL pilots' needs for support and providing appropriate aids with context awareness.

To mitigate this gap, this study adopts the widely adopted Large Language Model (LLM), GPT-4, as it has demonstrated remarkable achievement in context-awareness and understanding human intentions [84, 85]. The data processing of LLM is summarised in Section 2.4, and the LTL-based framework is presented in Chapter 4.

2.3 Linear Temporal Logic (LTL)

This study utilizes Linear Temporal Logic (LTL) to address the explainability challenge in identifying the OOTL status. LTL, a formal system in computer science, is used to specify and verify system behaviours over time. It encodes atomic propositions into a sequence of states and checks their compliance with LTL formulas, making it suitable for expressing the temporal logic of human behaviours [39, 86].

More specifically, the use of linear temporal logic on finite traces (LTL_f) formulas is proposed in this work to characterize behaviours from observed finite traces [87]. The LTL_f is a variant that extends the classical LTL to accommodate finite traces, enhancing its applicability to mine the temporal logic specification of system behaviours from a set of program execution logs in practical contexts [88, 89]. These execution logs, composed of traces that comprise a series of system states[90], are interpreted over finite traces using LTL_f formulas [91]. Therefore, the LTL_f is expected to process the finite trace of visual behaviours for identifying the OOTL status of pilots within a certain period.

2.4 Processing psychological data with Large Language Model (LLM)

The development of LLMs, such as GPT-4, has advanced data pattern analysis across various domains, particularly where large, complex datasets are involved. It offers a novel option for processing processing psychological data, including fNIRS [92], EEG [93, 94], and eye-tracking data [95, 96].

A recent study have demonstrated the application of LLMs in analyzing fNIRS-derived metrics to generate easy-to-understand evaluation reports and actionable recommendations for Mild Cognitive Impairment (MCI) rehabilitation tasks [92]. Their method combined the advantages of LLMs' pattern recognition capability to evaluate the MCI status of the subjects and LLMs' text generation ability to provide optimization recommendations for rehabilitation task design. Similarly, another study utilized a lightweight LLM in a local setting to recognize emotional states from EEG signals, generate personalized diagnostic and treatment suggestions, and automatically produce electronic medical records [94]. The adoption of LLMs in EEG signal processing provides transparent reasoning steps and enables interpretable step-by-step verification, thereby promoting trustworthiness in clinical contexts [93].

Compared to EEG and fNIRS, which are used to understand cognitive status at a general level, eye-tracking data has the potential to provide specific mappings to external stimuli in the environment. Monitoring eye movements with LLMs helps understand visual attention and information acquisition. For example, researchers developed a method to assist visually impaired individuals using a wearable eye tracker [96]. In this approach, the LLM recognizes points of interest from eye-tracking data and generates voice descriptions of objects, supporting users in better understanding their surroundings. However, the LLM in this work only reacts to the current visual attention, utilizing image-to-text and text-to-voice capabilities [97] to describe external stimuli. It lacks the ability to extensively analyse eye movements and discern the users' visual scanning behaviours.

To identify the needs for support of pilots, such as discerning troubleshooting behaviour from normal monitoring during flight, the LLM requires a broader understanding of pilots' eye movements. Therefore, a challenge lays in enabling the LLM to understand the pilot's visual attention in a certain period with inherent information processing logic. An appropriate pre-processing method for the eye-tracking data is needed to integrate with LLMs.

Chapter 3

Study 1: A Visual attention analysis model

In this chapter, a Flashlight model is proposed to enhance the granularity of eye-tracking data analysis. This model integrates both the attention distribution perspective and attention resource perspective of eye-tracking measurements to obtain a more comprehensive understanding of pilots' visual behaviours. The structure of this chapter is organized as follows: Section 3.1 presents the background with current studies, and proposes three research questions. Section 3.2 delves into the methodology for analyzing eye-tracking data utilizing the proposed Flashlight model in response to RQ1. Sections 3.3 and 3.4 showcase a case study and discuss the findings, addressing research questions RQ2 and RQ3, respectively. Finally, Section 3.5 summarises this work.

3.1 Background

Human performance degradation has long been acknowledged as a primary cause of aviation accidents [1], with over 70% of these accidents attributed to human factors such as fatigue, mind wandering, and diminished situational awareness [2, 3]. Especially with the fast development of automation systems, the inadequate workload in highly automated human-machine collaboration conditions may diminish the vigilance and situation awareness of the pilots and cause mind wandering, resulting in impaired human performance [82, 98, 99]. Therefore, monitoring the pilots' behaviours and predicting their performance has been attributed with significant importance in conducting in-time interventions and improving aviation safety [11, 100].

As introduced in Section 2.2.1, the development of performance prediction methods based on eye-tracking data has emerged as a prominent research direction aimed at risk prevention and safety enhancement in the aviation industry [33, 101]. However, Section 2.2.2.1 presents a research gap in the more comprehensive understanding of pilots visual attention, due to negelecting the integration of attention distribution and attention resource perspectives. To mitigate the research gap, we aim to explore three research questions (RQs) as following:

- RQ1: How to develop eye-tracking metrics that can reveal both the attention distribution and attention resource perspectives?
- RQ2: Are the developed metrics effective in predicting pilot performance?
- RQ3: How do the developed metrics contribute to pilot performance prediction?

In response to the first research question (RQ1), this study introduces an innovative Flashlight model for eye-tracking analysis that integrates two critical perspectives. As shown in Figure 3.1, the model combines the concept of William James' Spotlight attention model [102] with the widely adopted Christopher Wickens' model of human information processing [103]. Within this Flashlight model, the "bulb" symbolizes the focus of (visual) attention distribution, while the "battery" signifies the available attention resources for processing perceived information. This integrative model provides an intuitive demonstration that each engagement with an AOI demands the attention resources for supporting the entire information processing cycle. In response to RQ2 and RQ3, a case study is conducted to evaluate the AOI-gaze metrics developed based on the Flashlight model. It assesses the efficacy of the developed metrics and compares them with established input metrics to demonstrate how they contribute to pilot performance prediction.

3.2 Flashlight model

This section elucidates the rationale behind the synthesized flashlight model and explicates the utilization of eye-tracking metrics through an integrated approach.

3.2.1 Visual attention metrics in human performance

Human performance depends on both the attention distribution and the concurrent availability of attention resources. A review of studies in road traffic showed that though drivers can fairly handle the tracking task while talking on cellphones, driving performance was largely impacted when dialling [104]. Driving while talking on the phone can be successful multitasking since it uses different perceptual

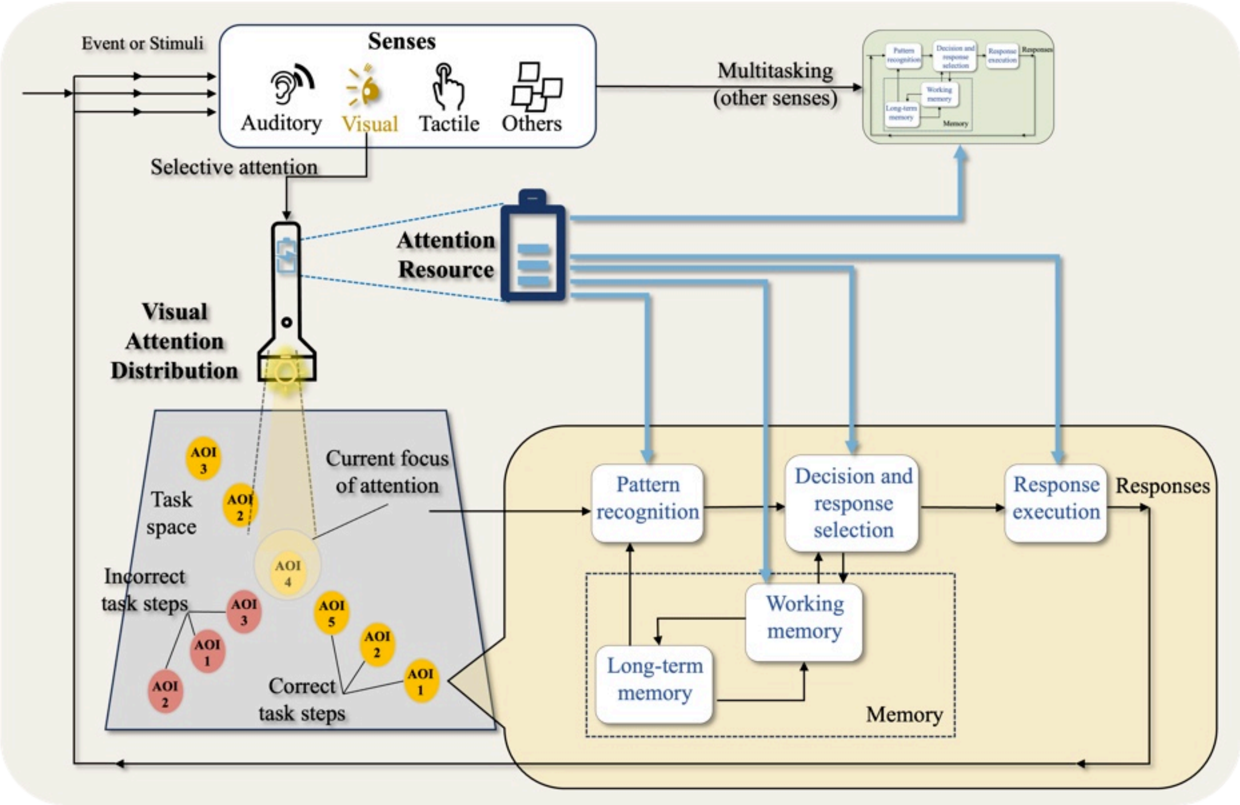

Figure 3.1: Flashlight model for visual attention

modalities (visual and auditory) in parallel, but driving while dialling requires visual selective attention sequentially [28]. Therefore, the decrement in performance is attributed to the requirement of focal vision for both tasks, necessitating directed fixations on the AOIs for sufficient information gathering [105, 106]. Given the limited field of view (approximately 4 degrees), an individual's focal vision is restricted to a singular AOI at any given time. This limitation suggests that the AOI which occupies a larger percentage of time is more likely to provide more vital information, making visual attention distribution critical for performance when tasks require selective attention sequentially [42]. Additionally, both tracking and dialling tasks demand attention resources for a complete information processing cycle, from pattern recognition to response execution [107]. The extent of available attention resources determines the capacity for processing information and, consequently, human performance.

3.2.2 Development of Flashlight model

This work proposes the Flashlight model by synthesizing the attention distribution and the attention resource perspectives as presented in Figure 3.1. The model provides an illustration for tasks requiring sequential visual selective attention. As shown in the upper part of Figure 3.1, the multitasking cases requiring divided attention (typically using different senses) is not discussed in the Flashlight model.

The Flashlight model integrates elements from James' Spotlight attention model on its left side, emphasizing how attention distribution across various tasks impacts performance [102]. Each circle on the task space indicates an AOI visit or revisit behaviour in sequence when handling the given task. Notably, the completion of a task might require acquiring comprehensive information from multiple AOIs or revisits to the same AOI. In this way, the flashlight can be deemed as a filter for visual attention, which decides what information is captured and processed. The correct task steps in yellow represent the optimal sequence of AOI visits for acquiring necessary information, and the incorrect task steps in red represent the AOI visits in a less effective way. Specifically, performance decreases when attention is unexpectedly directed towards incorrect paths (i.e., AOIs unrelated to the current task, or collecting information with incorrect sequence).

Meanwhile, the model incorporates Wickens' framework to outline the attention resources demanded for information processing [103]. In the complex and dynamic environment of aviation, pilots are tasked with continuously monitoring both internal instruments and external views, gathering data and making decisions across a spectrum of AOIs [108, 109]. Each engagement with the AOIs, represented as a circle on the left side, necessitates attention resources to support information processing. As a result of information processing, the decisions or the responses will direct the next step of selective attention for perceiving stimuli or conducting operational movements. This Flashlight model combines the attention distribution perspective and attention resource perspective, demonstrating their mutual effect on human performance.

3.2.3 Eye-tracking metrics based on Flashlight model

Eye-tracking metrics provide a real-time and unobtrusive method to monitor the two perspectives in the Flashlight model simultaneously. On the one hand, attention distribution metrics, such as total fixation time, total number of fixations, and visiting sequence, directly mirror how pilots allocate their attention and the effort expended in extracting information from visual stimuli [40, 110]. On the other hand, attention resource metrics are employed in various studies to infer information processing capability, reflecting factors like fatigue, situational awareness, and workload [33, 111].

Building on the demonstrated efficacy of eye-tracking metrics in predicting human performance from dual perspectives, this work advocates for an integrated approach of AOIs and attention resource metrics to facilitate a more thorough analysis. Given that information processing effort is inherently linked to each visit to an AOI, the eye-tracking measurements that are often used to gauge pilots' total available attention resources might be different on different AOIs. For example, though the pupil diameter is commonly used to monitor the pilots' cognitive status during tasks and predict task performances at an overall level [27, 70], the result from a study shows that there can be a significant difference between the pupil diameter on different AOIs in the same phase [76]. It demonstrate the necessity for assessing the attention resources devoted to specific AOIs, rather than at an overall level, to increase the granularity of pilots' attention analysis.

Following the Flashlight model's premise that *each engagement with an AOI demands the attention resources for supporting information processing*, this work develops the combined AOI-gaze metrics, as depicted in Figure 3.2 (Note that the AOI-gaze metric illustrated in the figure is exemplary and not exhaustive). These AOI-gaze metrics apply the measurements, that were typically used to measure total available attention resources, to individual AOIs to enrich the granularity of eye-tracking analysis. In addition to the attention distribution metrics that delineate how attention is allocated across various AOIs, and the attention resource metrics that measure overall information processing capacity, the AOI-gaze metrics offer a complementary approach to precisely evaluate the attention resources allocated for processing specific information from individual AOIs.

To validate the effectiveness of the AOI-gaze metrics and how they contribute to pilot performance prediction with other metrics, a case study is presented in Sections 3 and 4.

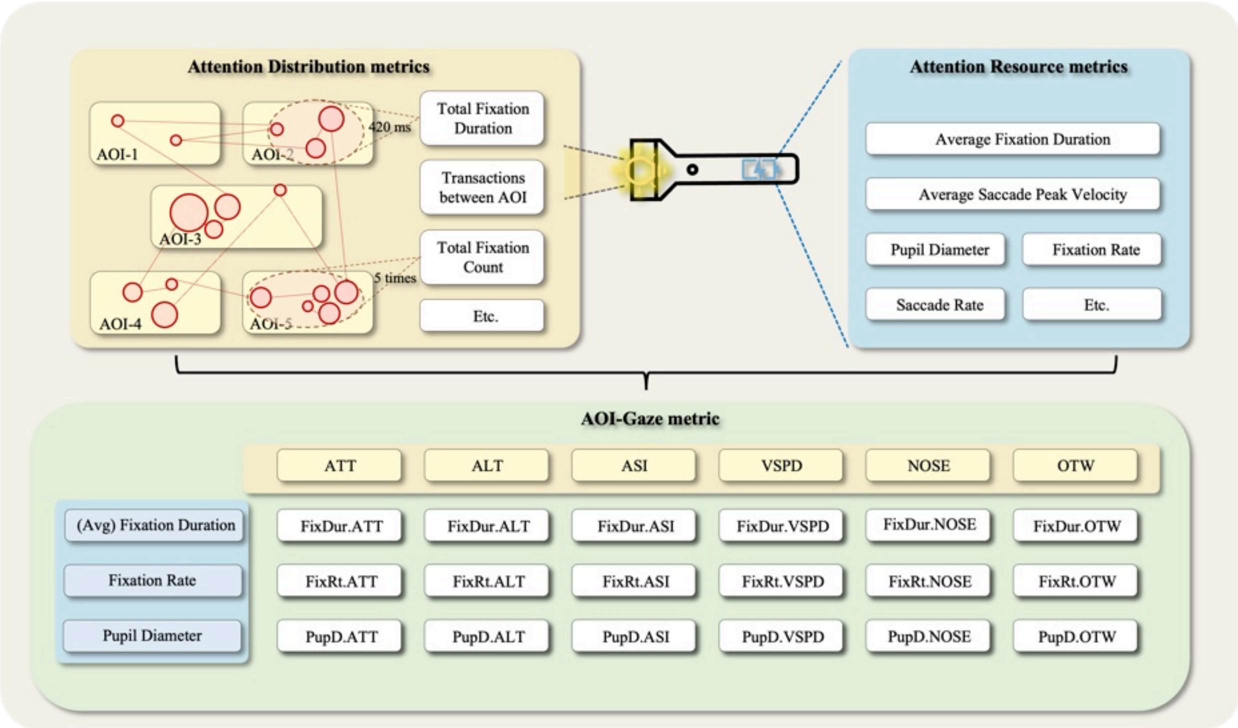

Figure 3.2: Combined AOI-gaze metrics

3.3 Case study

This study evaluated the AOI-gaze metrics developed based on the proposed Flashlight model through a flight simulation experiment. This section details the experiment settings, data collection, and evaluates the efficacy of the proposed metrics. It also compares the various types of input metrics using the widely recognized Gradient-boosted decision trees (GBDT) model to identify the most influential eye-tracking measurements in flight control tasks and demonstrate how these metrics contribute to pilot performance prediction.

3.3.1 Participants and apparatus

The study involved twenty-six students, aged between 22 and 32 years (M = 24.9, SD = 2.01), all of whom had normal or corrected-to-normal vision. These participants were recruited from The Hong Kong Polytechnic University (PolyU), The Chinese University of Hong Kong, and the City University of Hong Kong. The group consisted of 17 men (65.4%) and 9 women (34.6%). Four of these participants had prior experience (less than 5 hours) with the simulator, while the remaining participants were novices. This research complied with the American Psychological Association Code of Ethics and received ethical approval from the PolyU Institutional Review Board (Reference number: HSEARS20211117002), and written informed consent was obtained from all subjects before the experiment commenced.

The research experiment was conducted in the Department of Aeronautical and Aviation Engineering at PolyU. The participants flew in a Cessna 172 simulator providing a variety of controls, including yoke, rudder, throttle, and flaps. The simulator also provides a panel with instruments, such as attitude indicator, altimeter, and airspeed indicator. The graphics were produced by Microsoft Flight Simulator X and displayed on an assembled screen. The flight information, including pitch, bank, airspeed, and altitude, was recorded at 25 Hz. Eye-tracking data were collected using a wearable eye tracker with a frequency of 100Hz (Tobii Pro Glasses 3). The eye tracker contained 16 illuminators and 4 eye cameras to capture eye movements and pupil measures, and a scene camera to record the view ahead. Additionally, a desktop computer with a 27-inch monitor (1920*1080 pixels) was used to monitor the flight simulation and record eye-tracking data. Figure

3.3 illustrates the apparatus and experiment settings.

Figure 3.3: Apparatus and experiment settings

3.3.2 Experiment procedure and task design

The experimental process commenced with an introductory session and a free practice round, allowing participants to familiarize themselves with the simulator's controls and environment. To ensure that all participants were capable of handling the flight control tasks and completing the flight, the experiment employed a simplified manual operation requirement under visual flight rules (VFR) [21]. Participants were only required to maintain a Straight-and-Level flight [112] by adjusting pitch, roll, and yaw angles using joysticks and throttle, guided by instrument panels and external views. Other tasks, such as navigation and fuel monitoring, were automated, requiring no effort from the participants. The practice sessions lasted until participants were able to independently execute a complete flight, approximately 25 minutes in duration, from takeoff at Shanghai Hongqiao International Airport (ZSSS) to landing at Shanghai Pudong International Airport (ZSPD) under the simplified VFR condition. After obtaining signed consent forms, each participant was equipped and calibrated with the eye tracker before commencing the official experiment flights. A break of 10 to 20 minutes, depending on the participants' needs, was provided after the first flight to mitigate the effects of fatigue on subsequent performance. The order of the two flight routes was randomized to further ensure experimental integrity.

In the experiment, each participant was tasked with flying two segments of a round trip between Hong Kong and Guangzhou, which were recorded as separate files within the flight simulator. For each session, one of these files was selected and initiated with the aircraft already at cruising altitude (2,800 feet) under the autopilot function. After approximately 42 minutes of flight, an engine shutdown scenario was uniformly introduced to evaluate the participant's ability to maintain control of the aircraft's attitude without engine power, as depicted in Figure 3.4. The second session employed the return leg of the first session's route to maintain consistent variables such as flight distance and external conditions. This method also aimed to minimize the learning effect by offering a change in the external views, thus reducing participants' familiarity with the route from the initial session. All experiment trials were conducted under the simplified VFR condition, and all the participants completed the two segments successfully. Participants were unaware of the timing of the engine shutdown, and their manual aircraft control data after the engine shutdown were recorded to assess their performance in emergency takeover situations.

3.3.3 Flight control performance

Participants' flight control performance was assessed by analyzing the aircraft's pitch and roll angles during the descent phase, as illustrated in Figure 3.5. Fol-

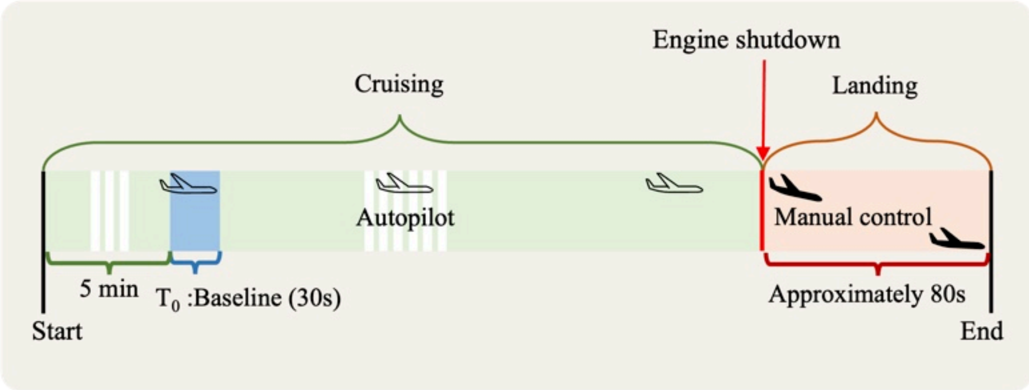

Figure 3.4: Experiment flight process

lowing the engine shutdown, both the altitude and airspeed of the aircraft started to decline due to the absence of power. The incident was triggered over open wilderness during the cruising phase, with no airports within view, to maximize the element of surprise for the participants. They were informed that their primary objective was to stabilize the aircraft (specifically, to minimize the pitch and roll angles) using the simulator's joysticks throughout the descent, rather than attempting an emergency landing on flat terrain. Consequently, participants were not tasked with locating a suitable landing site; their focus was solely on preserving the aircraft's attitude in mid-air. Specifically, though four participants who had prior experience, their performance didn't show significance comparing to the other participants who had only practice before the experiment.

Equations 3.1 and 3.2 define the Average Pitch (A_p) and Average Roll (A_r) to evaluate the pitch and roll angles during the descent phase at an overall level (t_s: time of engine shutdown, start of descent; t_g: time of ground contact). The absolute value of the pitch and roll angle (p_i and r_i) at each time point, multiplied by the time interval, indicates the proportion of the green (red) area shown in Figure 3.5. Lower values of A_p and A_r indicate better performance, with the aircraft maintaining a flatter attitude during the descent. A geometric mean of these two angles was calculated to provide a composite measure of flight control performance (M_{Geo}), which is obtained by taking the square root of the product of A_p and A_r. A lower M_{Geo} value denotes enhanced flight control performance, as articulated in Equation 3.3.

$$A_p = \frac{\Sigma_s^l |p_i| * \Delta t_i}{t_g - t_g} \tag{3.1}$$

$$A_r = \frac{\Sigma_s^l |r_i| * \Delta t_i}{t_g - t_g} \tag{3.2}$$

$$\Delta t = t_i - t_{i-1}$$

p_i: pitch value at sampling point i,$-180 < p_i < 180$

r_i: roll value at sampling point i,$-180 < r_i < 180$

$$M_{Geo} = \sqrt{A_p * A_r} \tag{3.3}$$

To better fit in the performance prediction model, this study conducted an inverted range-specific normalization to the M_{Geo} based on Min-Max method. Specifically, the M_{Geo} of each flight is normalized into the range [0.1, 0.9], considering there wasn't a ”perfect” or ”worst” performance. And interpretation of the M_{Geo} was inverted to be more intuitive: the higher value suggests better flight control performance. The normalization method is given in Equation 3.4.

$$(Normalized)M_{Geo} = 1 - (0.8 * \frac{M_{Geo} - Min}{Max - Min} + 0.1) \tag{3.4}$$

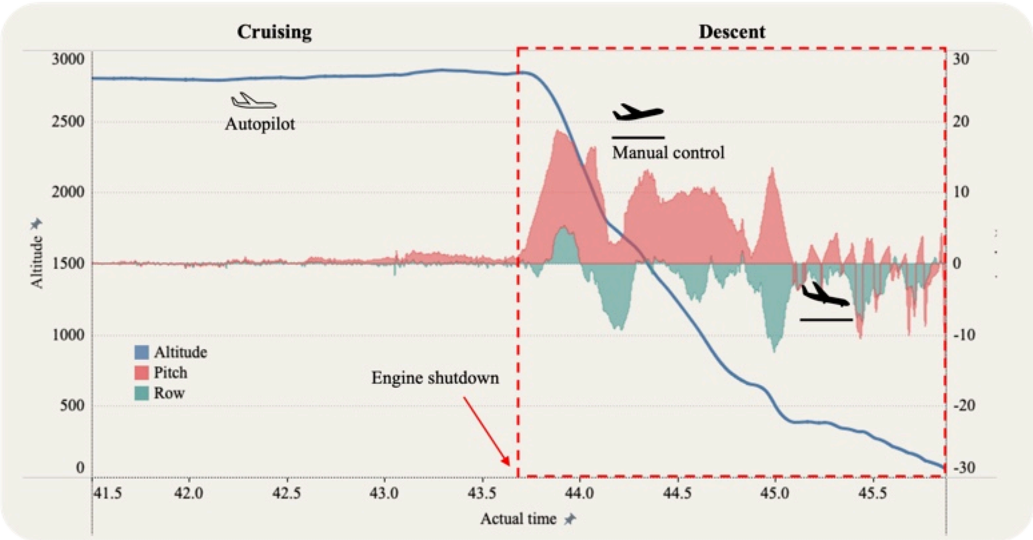

Figure 3.5: Pitch and row angle during descent

3.3.4 Eye-tracking measurements

In this study, eye-tracking data were processed using the *Tobii Pro Lab* software, employing three types of data depicted in the Flashlight model.

Attention Distribution Metrics: Six Areas Of Interest (AOIs) critical for monitoring aircraft status and managing flight control tasks were identified, as illustrated in Figure 3.6. These AOIs include essential instruments and visual cues within the flight simulator: the attitude indicator (ATT), altimeter (ALT), airspeed indicator (ASI), vertical speed indicator (VSPD), the aircraft's nose (NOSE), and the view outside the window (OTW). The ATT, ALT, ASI, and VSPD are internal instruments that provide critical information on the aircraft's status, whereas the NOSE and OTW provide a direct visual representation of the external environment and the aircraft's orientation relative to the horizon. *Dwell time(%)*, encompassing both fixations and saccades within these AOIs, is calculated to represent the pilots' visual attention distribution during the task. For instance, *Dwell (%).ATT* quantifies the percentage of total fixation and saccade time dedicated to the attitude indicator during the descent phase.

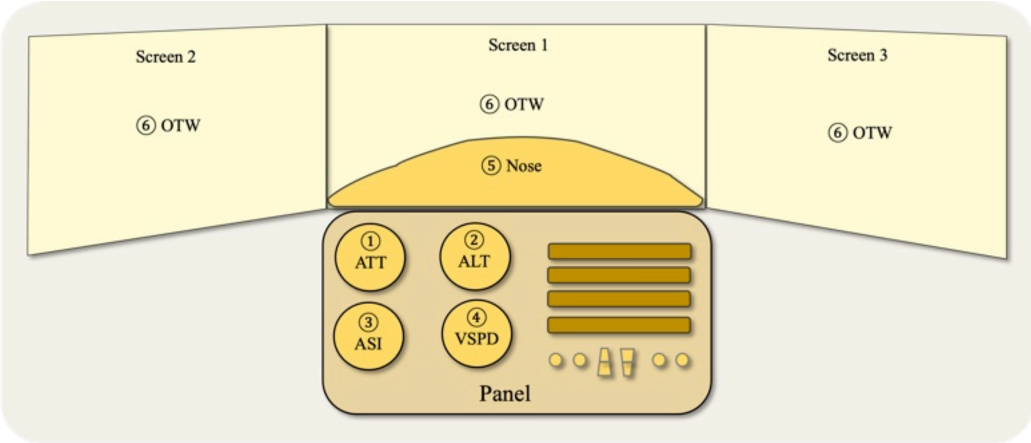

Figure 3.6: Areas of Interest (AOIs) in the flight simulator

Attention Resource Metrics: This analysis included five metrics related to attention resources: Pupil Diameter (*PupD*) [113, 114], Average Fixation Duration (*FixDur*) [115, 116], Fixation Rate (*FixRt*) [117, 118], Saccade Rate (*SacRt*) [119, 120], and Saccadic Peak Velocity (*SacPkV*) [121, 122]. *PupD* was analyzed using a baseline normalization method to account for individual differences in pupil size [123]. A thirty-second baseline period (T_0), starting five minutes into each trial, was established to capture the subjects' normal working state, as shown in Figure 3.4. The *PupD* during the descent phase was then normalized against the T_0 period to indicate relative pupil size changes. *FixDur* and *FixRt* measure, respectively, the average duration of fixations and the average number of fixations per second during the descent. *SacRt* and *SacPkV* calculate the average number of saccades per second and the average peak velocity of these saccades during the same phase.

AOI-gaze Metrics: The study merged the six AOIs with three attention resource metrics (*FixDur*, *FixRt*, *PupD*), resulting in 18 combined AOI-gaze metrics, as depicted in Figure 3.2. For example, *FixDur.ASI* is the average fixation duration on the airspeed indicator, calculated by dividing the total fixation duration on ASI by the total number of fixations on ASI. *FixRt.ASI* reflects the fixation rate on ASI, obtained by dividing the total number of fixations on ASI by the total descent phase duration. *PupD.ASI* represents the average pupil diameter when viewing ASI during the descent, normalized against the global pupil diameter during T_0. Saccade metrics were excluded from this analysis due to the minimal occurrence of saccades within the instrument AOIs.

Mixed Metrics: To investigate how the developed AOI-gaze metrics contribute to pilots' performance with other metrics, all the eye-tracking measurements from the previous three categories are incorporated into one group. It serves the purposes of validating if the inclusive analysis of all three categories can exceed singular usage and identifying the most pivotal eye-tracking measurements across all the categories.

3.3.5 Data analysis and comparison

To explore RQ2 and RQ3, this study utilized the eye-tracking metrics outlined in Section 3.3.4, applying them within the framework of the widely adopted Gradient-boosted Decision Trees (GBDT) model for performance prediction [124, 125]. The analytical process is represented in Figure 3.7.

Initially, the metrics were categorized into four distinct groups for analysis: (1) Attention Distribution Metrics, (2) Attention Resource Metrics, (3) AOI-gaze Metrics, and (4) Mixed Metrics. Subsequently, these metrics were integrated into the GBDT model to facilitate feature selection.

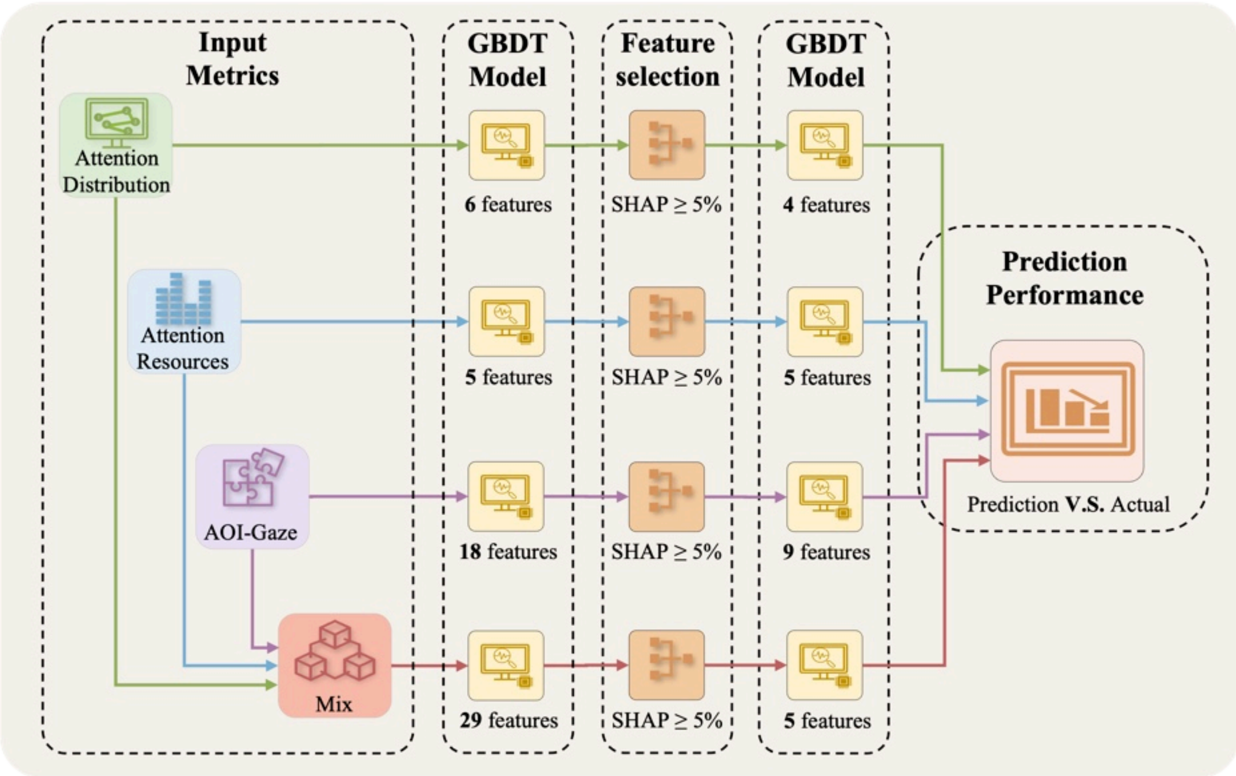

Figure 3.7: Data analysis procedure

A key part of this process involved calculating the SHapley Additive exPlanations (SHAP) values to identify the measurements most contributory to model output [126]. SHAP value is based on game theory and offers a comprehensive method for assessing feature importance that is applicable across various machine learning models, including both linear and nonlinear frameworks. These values account for feature interactions and dependencies, providing a detailed explanation of model predictions. The SHAP value for a given feature i in a prediction $f(x)$ represents the change in the expected model output attributable to the feature's observed value x_i. Features with higher SHAP values are deemed to have a more substantial effect on the prediction model, suggesting a stronger linkage between the specific eye-tracking measurement and flight control performance. The average SHAP value for the jth feature is calculated using the following formula:

$$\sum_{j=1}^{p} \phi_j(\hat{f}\,) = \sum_{j=1}^{p} |\theta_j x_j - E(\theta_j X_j)| \tag{3.5}$$

where p represents the total number of features and $E(\theta_j X_j)$ denotes the mean effect for feature j. Taking the absolute value allows for aggregating the total influence of each feature across all instances, irrespective of the direction of influence. Without this step, a feature with high impact in both positive and negative directions might be incorrectly neutralized.

Following the computation of SHAP values, the contributions of measurements within each metric group were totalled, and their proportional impact was deduced. Metrics contributing more than 5% to the total SHAP value within their respective groups were identified as selected features for further analysis. These highlighted metrics signify the most impactful eye-tracking measurements within each category and were subsequently reapplied to the GBDT model. To assess and compare the effectiveness of these metrics, the residuals of the predicted results versus actual performance (Normalized M_{Geo}) were computed and visualized. The subsequent section will detail the metrics identified and their comparative performance analysis.

3.4 Results

3.4.1 Performance of different metrics

The feature selection results based on SHAP values are presented in Table 3.1. Within the Attention Distribution metrics, it was found that the total dwell times on the airspeed and vertical speed indicators were excluded, whereas all initial metrics under the Attention Resource category were preserved. The AOI-gaze metrics distinguished themselves by having the highest number of measures with contributions surpassing the 5% threshold. Significantly, features pertaining to the attitude indicator (ATT) and airspeed indicator (ASI) demonstrated contributions above 5%. Moreover, the top three most impactful measurements within the AOI-gaze metrics were all associated with ATT. In the combined (Mix) group, metrics from Attention Distribution (i.e., *Dwell (%).NOSE* and *Dwell (%).ATT*), Attention Resources (i.e., *SacPkV*), and AOI-gaze (i.e., *FixRt.ATT* and *PupD.ATT*) were screened out.

Table 3.1: Top 5% measurements in each input group

Attention Distribution	Attention Resource	AOI-gaze	**Mix**
Dwell (%).NOSE	PupD	FixDur.ATT	Dwell (%).NOSE
Dwell (%).ATT	SacPkV	PupD.ATT	Dwell (%).ATT
Dwell (%).OTW	SacRt	FixRt.ATT	FixRt.ATT
Dwell (%).ALT	FixRt	FixRt.ASI	SacPkV
	FixDur	FixDur.ASI	PupD.ATT
		FixRt.OTW	
		FixDur.ALT	
		FixRt.ALT	
		PupD.ASI	

We evaluated the predictive accuracy of each metric set using 10-fold cross-validation, where the dataset is randomly divided into 10 equal parts, with nine parts used for training and one for testing in each of 10 iterations, ensuring robust model evaluation. The overall Mean Absolute Errors (MAE) of the 10 rounds were calculated from the residuals of the predicted versus actual performance, using Normalized M_{Geo} as a reference. The same dataset and 10-fold separation is adopted for all four metric sets to ensure the reliability of comparison results. As shown in Figure 3.8, AOI-gaze metrics displayed an accuracy (MAE = 0.122) comparable to that of Attention Distribution metrics (MAE = 0.123) and superior to Attention Resource metrics (MAE = 0.140), highlighting their predictive effectiveness. The mixed metrics, encompassing eye-tracking measurements from all categories, exhibited the highest predictive accuracy (MAE = 0.108).

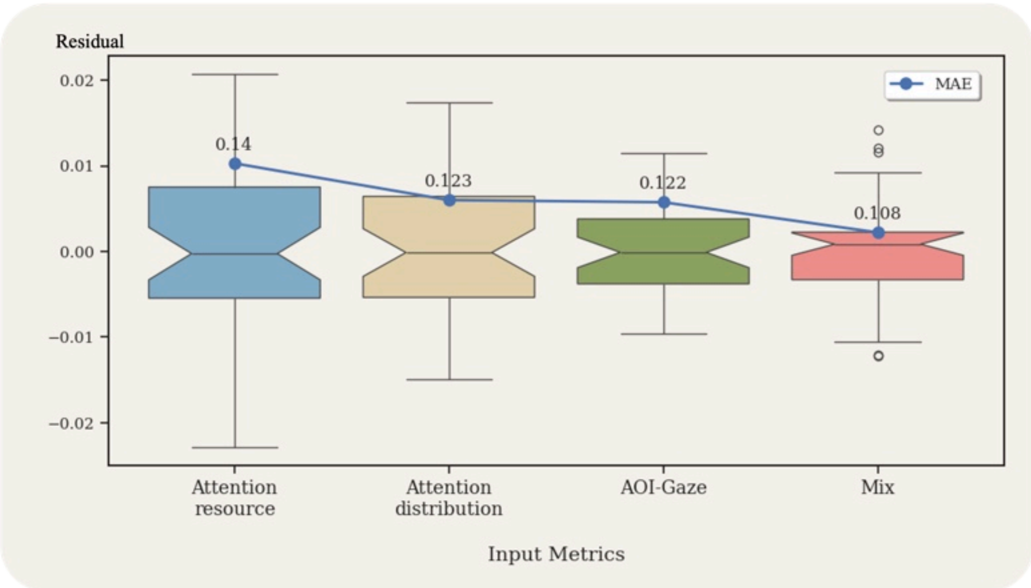

Figure 3.8: Prediction accuracy with selected features

Beyond assessing overall prediction accuracy through MAEs, we also examined the variance in prediction accuracy across different sections by plotting the results against the normalized M_{Geo}. Given the minimal instances (three) where performance was below 0.5 after normalization, only results with M_{Geo} greater than 0.5 were visualized. The AOI-gaze and Mixed metrics demonstrated more consistent accuracy with fewer deviations compared to the Attention Resource and Attention Distribution metrics. A noticeable peak was observed for AOI-gaze, Attention Distribution, and Attention Resource metrics at an M_{Geo} around 0.75, whereas the Mixed metrics exhibited less variation around this value, suggesting a more stable prediction across a range of performance levels.

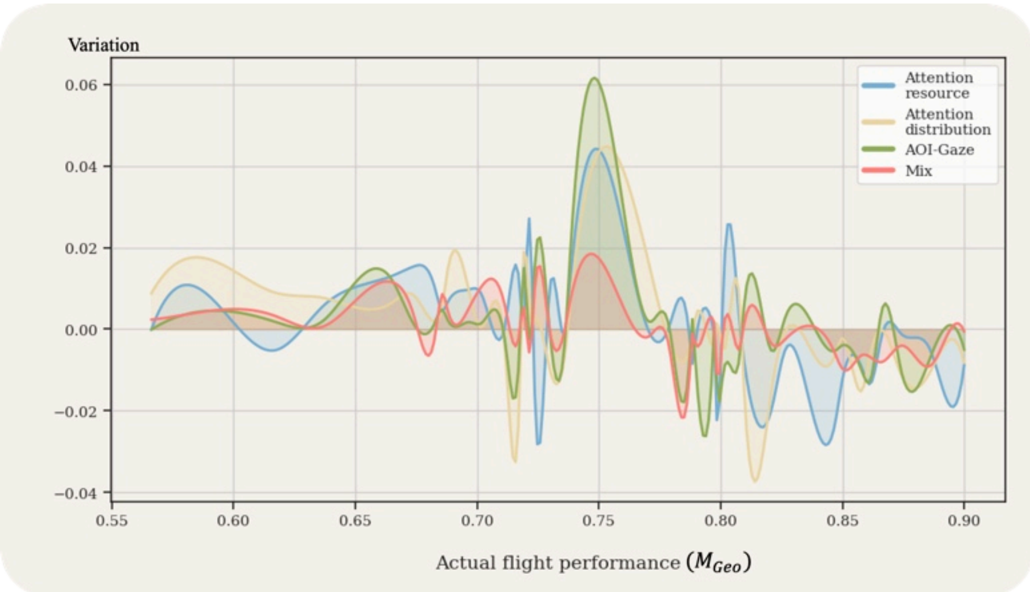

Figure 3.9: Variation of prediction results from actual performance

3.4.2 Discussions

The case study demonstrates the efficacy of combining different types of eye tracker metrics in analyzing the pilots' visual behaviours and predicting operation performance. By applying various input metrics within the Gradient-boosted decision trees (GBDT) models, the study illustrates the superior predictive accuracy of AOI-gaze metrics over singular applications of Attention Distribution and Attention Resource metrics. Subsequently, the three groups of input metrics were assembled together as Mix metrics to identify the most influential eye-tracking measurements using SHAP values.

3.4.2.1 Explanations of the results

In response to RQ1, the Flashlight model provides a systematic approach to analyse eye-tracking metrics from both the attention distribution and attention resource perspectives, highlighting the comprehensiveness of pilots' visual behaviours and human performance in complex and dynamic aviation tasks.

In response to RQ2, Figure 3.8 and Figure 3.9 demonstrated that the developed AOI-gaze metrics reached better performance than solely using the Attention distribution metrics and Attention resource metrics.

In response to RQ3, the results suggest that the inclusive usage of eye-tracking measurements from three categories achieved optimal prediction accuracy with better adaptation around potential outliers. It demonstrated the contribution of proposed AOI-gaze metrics in complementing the gap between attention distribution and attention resource metrics.

Specifically, five key measurements were identified in the flight attitude maintaining tasks: (1) the percentage of total dwell time on the nose of the aircraft (*Dwell(%).NOSE*); (2) the percentage of total dwell time on the attitude indicator (*Dwell(%).ATT*); (3) the fixation rate on the attitude indicator (*FixRt.ATT*) (4) the average saccade peak velocity across the whole scenario (*SacPkV*); (5)

the average pupil diameter of fixations on the attitude indicator (*PupD.ATT*). The *Dwell(%).NOSE* suggests the importance of the aircraft nose as it offers a direct visual representation of the aircraft's orientation relative to the horizon. The *SacPkV* indicates the importance of active saccade activities to the flight control performance, in line with the previous research revealing that higher saccade velocities imply lower fatigue level and higher vigilance [67, 68]. The *Dwell(%).ATT*, *PupD.ATT*, and *FixRt.ATT* demonstrate the importance of the attitude indicator in flight attitude maintenance tasks, in line with the finding of White et al.[25].

3.4.2.2 Implications

With the effectiveness shown in this case study, the proposed Flashlight model provides guidance for developing novel AOI-gaze metrics to enhance the granularity of the eye-tracking analysis using the information processing efforts on specific AOIs. Furthermore, the selected measurements from the Mix metrics obtained the best accuracy among the tested input metrics, suggesting a comprehensive approach to capturing the multifaceted nature of human behaviour and performance.

Meanwhile, the results provide deeper insights into gaze behaviour towards AOIs. For instance, though the highest SHAP value of *Dwell(%).NOSE* in the Attention distribution metrics suggests the importance of the total time distributed to the nose of the aircraft, either average fixation duration or average fixation rate on the nose (*FixDur.NOSE* and *FixRt.NOSE*) didn't contribute more than 5% in the AOI-gaze metrics. It potentially implies that while the total dwell time on the nose is critical to maintaining the attitude of the aircraft, the manner of monitoring (e.g., duration or frequency of fixations) may vary in importance.

Finally, though previous studies have revealed that the pupil diameter, average fixation duration, and fixation rate can reflect the pilots' cognitive status and be used in predicting pilots' performance [31, 33], this study further revealed that the specific AOI-gaze metrics can be more effective in prediction performance with an enhanced granularity.

3.4.2.3 Summary

In summary, the proposed Flashlight model provides a systematic basis for analysing human performance with a combination of both attention distribution and attention resource perspectives. Following this model, the developed AOI-gaze metrics serve as an effective complementary to the conventional Attention distribution metrics and Attention resource metrics in predicting pilots' performance. The most influential eye-tracking measurements identified with SHAP values enhance the granularity of data analysis and provide more insights into the visual behaviours in flight control tasks. This work contributes to the pilots' performance prediction from the theoretical level to the specific measurement level, promoting the comprehensive analysis of pilots' behaviour and prediction of pilots' performance. However, the proposed Flashlight model only discusses the tasks using sequential selective attention. More explorations can be made by incorporating the tasks requiring parallel selective attention and divided attention, to further enhance the understanding of human attention.

3.5 Concluding remarks

Predicting pilots' performance is crucial for timely interventions and improving aviation safety. Eye-tracking metrics have become instrumental in analyzing pilots' behaviours and predicting operational performance within complex and dynamic human-machine interaction environments. Nonetheless, more studies are needed on the comprehensive integration of attention distribution metrics and attention resource metrics. To bridge this gap, our study introduced the Flashlight model to explain the pilots' visual behaviours, integrating James' Spotlight attention model and Wickens' information processing model.

Our work facilitates a more comprehensive analysis of eye-tracking data and the development of pilots' performance prediction methods with two significant contributions. Firstly, it innovates by developing combined AOI-gaze metrics that advance the analysis of information processing towards specific Areas of Interest (AOIs), thereby refining the detail available in gaze metrics analysis. This approach offers a fresh perspective for pinpointing critical AOIs in performance prediction beyond mere total fixation or dwell durations. Secondly, the study employs SHAP values to highlight the impact of various eye-tracking measurements on the model predicting performance. Through this method, it identifies the most pivotal eye-tracking measurements for predicting performance in flight attitude control tasks, enriching the understanding and application of eye-tracking data in aviation research. The model and methods proposed in this study could be used in pilots' training to conduct a more comprehensive analysis of the pilots' performance based on eye-tracking techniques, and provide more detailed instructions on how the pilots could improve their visual scanning behaviors for better task performance.

While our study contributes several insights into pilots' visual behaviours and performance prediction using eye-tracking metrics, it is not without its limitations. The inclusion of student participants and the utilization of simplified flight control tasks could potentially limit the applicability of our findings to more complex and realistic flight scenarios. Additionally, the study did not extensively explore certain eye-tracking metrics, such as blinks and gaze entropy, which may offer further insights into pilots' attention and cognitive states. Future research is planned to address these limitations by enhancing the realism of the experimental setup, involving licensed pilots, and incorporating more intricate flight tasks. This will not only improve the validity of our findings but also allow for a more detailed examination of pilots' behaviours under varied and challenging conditions. Furthermore, a broader range of eye-tracking metrics will be examined in depth, facilitating a more comprehensive analysis of pilots' visual behaviours and improving the accuracy of performance predictions. Through these advancements, we aim to deepen our understanding of pilot behaviour in complex and dynamic human-machine interaction environments, ultimately contributing to safer and more efficient aviation operations.

Chapter 4

Study 2: A explainable OOTL identification framework

This chapter introduces a framework to identify pilots' OOTL status with explainability. The framework innovatively leverages the Linear Temporal Logic (LTL) methods to obtain human-readable formulas for an explainable identification result, as well as more insights into human behaviours. This chapter is organized as follows: Section 4.1 provides the background and motivation of this work by reviewing the characteristics of OOTL phenomenon and identifying the challenge of applying LTL methods. Section 4.2 details the framework of the proposed VALIO methods. Section 4.3 presents a case study to validate the proposed approach and discusses the results in comparison with other benchmark methods. Finally, Section 4.4 outlines the main contributions and limitations of this work and highlights future research directions.

4.1 Background

4.1.1 Identifying the OOTL status

The introduction of advanced autopilot systems and flight management functions has significantly mitigated human error in aviation by taking over manual flight operations traditionally executed by pilots [5]. These systems, known as Autopilots and Flight Management Systems (FMS), automate a multitude of flight operations, including aircraft control, navigation, information display, and fuel management. Automation reduces pilot workload and the probability of human-induced aviation accidents [127]. However, an unintended consequence of increasing reliance on automation is the potential for pilots to become progressively disengaged from the control loop. Studies suggest that prolonged exposure to high levels of automation (LOA) can lead to decreased focus, vigilance, and situational awareness among pilots, thereby increasing fatigue and reducing skill proficiency [61, 128, 129]. This detrimental effect on human performance, resulting from the absence of active human involvement in the control loop, is known as the Out-Of-The-Loop (OOTL) phenomenon [10]. Data from both NASA and a survey among German aviators indicate that the OOTL phenomenon significantly contributes to human errors in aviation [12, 13].

Characterizing and quantifying the OOTL phenomenon remains a challenge, as it is not confined to a specific domain but manifest across several factors in the information processing tasks [130]. For example, vigilance failure has been identified as a critical factor of OOTL phenomenon [131, 132]. Additionally, mind-wandering (MW) has emerged as a significant aspect in the examination of OOTL, further expanding the scope of research in this area [44]. And the work of Merat et al. (2019) is notable for delineating three specific statuses,"in-the-loop, on-the-loop, and out-of-the-loop", based on situation awareness (SA) within the context of automobile monitoring [8]. Besides, the discussion has also been broadened with other related factors such as daydreaming and distraction, underscoring the complexity of OOTL phenomenon[133, 134]. To investigate these factors, researchers have turned to biometric measurements for objective psychophysiological data, utilizing tools like Electroencephalograms (EEG) and Functional Near-Infrared Spectroscopy (fNIRS) to track vigilance degradation due to passive fatigue, or employing pupil diameter and saccade measurements to identify OOTL instances triggered by mind wandering [22, 34, 135–137].

While the problem of being OOTL has been acknowledged and extensively studied from diverse perspectives for several decades, there remains a significant gap in the explainability of methods used to identify OOTL status [8, 44]. As introduced in Section 2.2.2.2, the inability to understand the reasoning behind decisions and the potential for errors could have severe consequences [35, 36]. This opacity of "balck-box" methodologies poses a significant challenge in aviation operations, as inaccurately identifying OOTL status can trigger unwarranted alarms or interventions. Such false positives may contribute to alarm fatigue and detrimentally affect pilot performance [37]. Conversely, providing clear explanations for OOTL detections can enable the creation of more human-centred and warranted alerts that help pilots recognize and adjust their behaviours, thereby mitigating the negative impacts of OOTL status. This chapter presents a work aimed to develop an OOTL identification approach with explainability using Linear Temporal Logic (LTL) methods.

4.1.2 Linear Temporal Logic for behaviour classification

The LTL method, initially developed for the formal verification of computer programs, encodes atomic propositions (e.g., "The value of variable V2 has changed") in a linear sequence of states, or traces, to represent system behaviours [86, 138]. It checks whether these traces satisfy certain LTL formulas, constructed using a set of propositional variables, logical operators (negation and disjunction), and temporal modal operators (next and until). The combination of logical and temporal operators makes it particularly suitable for expressing the temporal logic of human behaviours. Moreover, the explicit formulas facilitate human understanding of the classification model and enhance the explainability of the results [39].

As introduced in Section 2.3, this study uses linear temporal logic on finite traces (LTL_f) formulas to characterize behaviours from observed finite traces [87]. A variety of methods have been developed to learn arbitrary LTL_f formulas. For instance, an alternating automaton is constructed by exploring a skeleton space to model observed behavior [39]. A Bayesian probabilistic model is also employed to infer contrastive explanations that delineate differences between various traces for learning LTL_f formulas [139]. Gaglione et al. proposed a method to infer minimal LTL formulas by transforming the inference problem into a maximum satisfiability (MaxSAT) problem and then utilizing off-the-shelf MaxSAT solvers to find a solution [140]. However, these approaches may pose challenges in the analysis of dynamic and complex eye-tracking data as they either presuppose a noise-free environment [39, 141], limit the hypothesis space by LTL_f template [86, 142], or are subject to the high complexity inherent in MaxSAT [140].

This work proposes the integration of Graph Neural Network (GNN) inference to support the inference of LTL_f, building on the work of Luo et al. [91]. Their work demonstrates that GNNs can capture the satisfaction relations of the LTL_f formulas and simulate the LTL_f inference to distinguish between positive and negative traces. By utilizing GNN inference, the search problem in discrete space can be transformed into a parameter learning problem in continuous space, a topic that has been extensively studied in recent years. The remaining challenge lies in how to encode the massive continuous gaze movements into the finite traces composed of atomic propositions that can be processed by the GNN-based LTL_f method.

4.1.3 Parse visual attention through eye-tracking

Dynamic gaze movements can be encoded by quantifying the visual attention distribution, which involves assigning collected eye movement data to specific regions of the visual scene, known as Areas Of Interest (AOIs). AOIs inherently contain spatial information and serve as a foundation for encoding the temporal relationship of eye movements [120, 143]. Researchers widely adopt AOI-based metrics to encode spatial information in attention-based studies, as they provide specific semantic information across the entire visual scene [29, 62, 144]. For example, Schnebelen et al. partitioned the driving scene into 13 AOIs based on the positions and functionalities in road traffic [49]. These AOI-based metrics were then used to study drivers' visual strategies in highly automated driving and estimate OOTL. Similarly, Li et al. utilized AOI-based metrics to assess the information processing of military pilots in fighter aircraft [145]. Furthermore, Haslbeck et al. found that pilots employing different visual scan strategies exhibited varying levels of information processing efficiency and flight performance by dividing the instruments into different AOIs based on their functionalities [146]. These studies highlight that appropriate AOI segmentation aids in linking eye movements to stimuli and provides an objective measure of visual attention, thereby facilitating the interpretation of the underlying logic. In practice, appropriate AOI segmentation largely depends on the functionalities of the stimuli. For instance, in automobile driving studies, the scene out of the front window can be divided into more than five AOIs according to the relative position of the road (i.e., Center, Left, Right, Up, Down), as they possess different characteristics and provide different information in driving [49]. On the other hand, in aviation, the scene outside the window provides only limited information to the pilots, while the various instruments in the cockpit provide heterogeneous critical information related to the flight. As a result, the entire view outside the window is usually defined as a single AOI, while multiple AOIs are defined in the cockpit based on the instruments [55]. Therefore, a comprehensive understanding of the functionalities of the stimuli and their relationship to the tasks is crucial for properly defining the AOIs for analyzing visual attention distribution.

4.1.4 Challenges of applying LTL methods

Though the AOI-based eye-tracking data contains more contextual information for explaining the visual behaviours, a gap remains in adopting the AOI-based eye movement measurements to the LTL_f methods. The eye movement data collected by eye trackers can be automatically mapped to each AOI with the support of advanced computer vision technologies, generating AOI-based metrics [147]. These metrics, such as visit times and average fixation duration, reflect the characteristics of the gaze movements toward certain stimuli during specific periods. It is necessary to extract individual events from such data to form the traces that can be parsed by LTL_f methods and achieve explainable OOTL identification.

Many studies worked on encoding the temporal-spatial information in pilots' visual attention, while they are not compatible with the LTL_f methods. For example, McClung and Kang developed a method to characterize the scanning patterns of the Air Trafic Controllers [148], while it relies on the complete observation of all aircrafts (AOIs) on the display. Other studies encode the dwells into vectors of fixated length [149] or calculate the transition matrices across the AOIs [146, 150], while these methods omit the duration of each dwell. Though these methods are effective in the specific tasks, an approach for encoding the temporal-spatial information in the gaze movements that is compatible with the LTL_f methods is needed.

4.2 Visual Attention LTL_f for Identifying OOTL (VALIO)

To tackle the compatibility challenge discussed in Section 4.1.4, we proposed a method to encode pilots' visual attention distribution data into formalized traces and developed a framework, named *Visual Attention* LTL_f *for Identifying OOTL (VALIO)* in this section.

The VALIO framework, as illustrated in Figure 4.1, executes in three phases. Initially, AOIs are defined to reflect the stimuli in the cockpit, providing a foundation to encode the spatial information of pilots' gaze movement. Following this, eye movement data collected via eye trackers are transformed into Visual Attention Traces (VAT). The AOIs being observed and the corresponding durations are encoded as atomic propositions in the traces, making the eye movement data compatible with the LTL methods. Subsequently, these VATs are employed to train a Graph Neural Network (GNN), yielding a set of parameters. These parameters are interpreted as LTL_f formulas, with the most effective formula selected through a strategic approach. Finally, insights into pilots' OOTL status can be obtained by interpreting the obtained LTL_f formulas and analysing the pilots behaviours. The details of each phase are further elaborated in the following subsections.

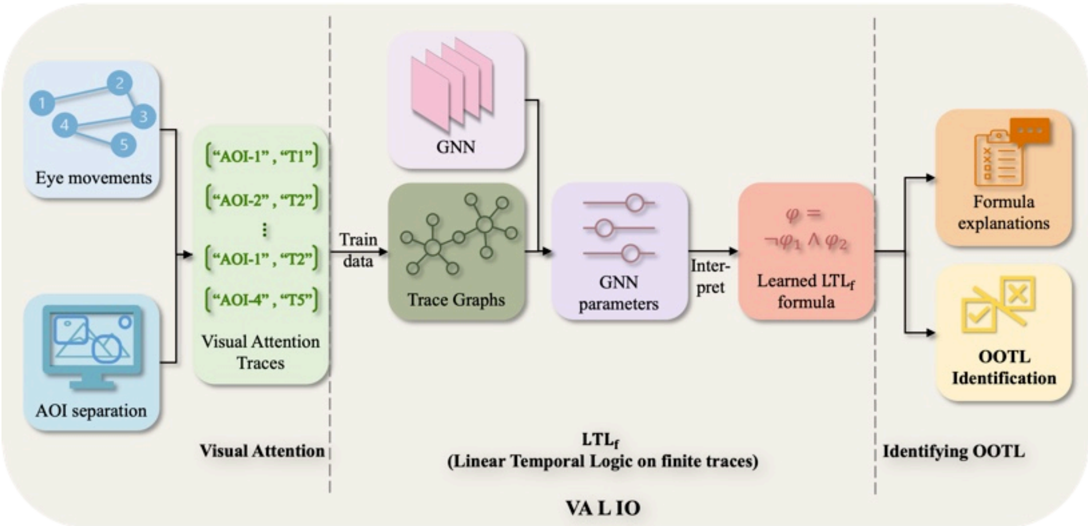

Figure 4.1: Framework of Visual Attention LTL_f for Identifying OOTL (VALIO)

4.2.1 Visual Attention Traces (VATs)

This research introduces a methodology to encode pilots' dynamic gaze patterns from eye-tracking data into Visual Attention Traces (VAT). VAT serves as a structured representation to analyze the effort involved in information acquisition through gaze behaviors, reflecting individual events of stimulus engagement [151].

Firstly, the proposed VALIO framework defines AOIs upon different instruments in the cockpits based on the distinctive designs of different aircraft. A more specific example is provided by the case study in Section 4.3. After defining AOIs in the cockpit, the Visual Attention Trace (VAT) is then generated based on the dwells. A dwell is defined as *the interval between a gaze entering an AOI and eventually leaving it, including all the fixations and saccades during this visit*, and dwell time can be associated with motivation and top-down attention [120, 152]. VALIO depicts the pilots' effort of acquiring information from an AOI by denoting a dwell $d_i(a_i, t_i) \in \mathcal{D}$ as equation 4.1 and 4.2:

$$a_i \in \{AOI_1, AOI_2, AOI_3, ...\} \tag{4.1}$$

$$t_i \in \{T_1, T_2, ..., T_\tau\} \tag{4.2}$$

in which $\mathcal{D}$ represents the set of dwells, i indicates the sequence of the dwell in the whole collected trace, a_i and t_i represent the AOI (instrument) being visited in this dwell and the dwell time. Notably, the continuous factor dwell time $t_i \in \mathcal{T}$ is binned into several levels in VALIO to encode it into an atomic proposition format that can be processed by the LTL_f methods. The division of the different time levels can be determined according to the specific tasks and the distribution of dwell time. The case study in Section 4.3 presents four methods to determine the time levels and compare their performances. Correspondingly, a VAT (denoted as v) with finite length is defined in the form as $v = d_0, d_1, ..., d_n$.

The length of a VAT ($|v|$) is determined by a specified duration within the VALIO framework. For instance, if a duration is set at approximately 20 seconds, the handling of a dwell d_n at the conclusion of a current VAT (v_α) which encompasses the 20-second mark, is executed as follows: If more than half of d_n falls within the 20-second timeframe, it remains as a part of v_α, thereby extending the length of $|v_\alpha|$ to slightly exceed 20 seconds. Conversely, if the majority of d_n extends beyond the 20-second mark, it is allocated as the initial dwell of the subsequent VAT ($v_{\alpha+1}$), resulting in $|v_\alpha|$ being marginally shorter than 20 seconds. More specifically, the final dwell (d_{-1}) in the current VAT (v_α) is determined as shown in equation 4.3 below:

$$d_{-1} = \begin{cases} d_{n-1}, & |\,|d_0, ..., d_{n-1}| - 20sec\,| < \,|\,|d_0, ..., d_n| - 20sec\,| \\ d_n, & |\,|d_0, ..., d_{n-1}| - 20sec\,| > \,|\,|d_0, ..., d_n| - 20sec\,| \end{cases} \tag{4.3}$$

where $|d_0, ..., d_n|$ represents the total duration from the first dwell (d_0) to the nth dwell (d_n). And $|\,|d_0, ..., d_n| - 20sec\,|$ calculates the deviation between the total duration of the included dwells and the designated VAT length (20 seconds). The first branch indicates that including the d_n will lead to a larger deviation from 20 seconds than excluding it, so the d_{n-1} is determined to be the last dwell (d_{-1}) in the current VAT (v_α). Consequently, d_n is designated as the initial dwell (d_0) of the next VAT ($v_{\alpha+1}$), leading the $|v^\alpha|$ to be slightly less than 20 seconds in length. The second branch indicates that excluding the d_n will lead to a larger deviation from 20 seconds than including it, so d_n is determined to be the last dwell in the current VAT. Thisapproach ensures a precise division of attention data into segments, preventing a complete dwell from being split into two parts by the fixed VAT length.

4.2.2 Linear Temporal Logic on finite traces (LTL_f)

The VAT generated in the previous steps provides a compatible format for the Linear Temporal Logic on finite traces (LTL_f) analysis. Originally, a LTL_f trace can be represented as $\pi = s_0, s_1, ..., s_n$, where $s_t \in 2^{\mathcal{P}}$ is a state at time t. s_i represents a state composed of several atomic propositions $p \in \mathcal{P}$. For every state s_i, p holds if $p \in s_i$, or $\neg p$ holds otherwise. The LTL_f traces are label as postive and negative traces depending on the system status needed to be verified. And the VATs corresponding to the pilots' OOTL status are also labeled as positive for training purpose. As shown in Figure 4.2, the VAT in the previous steps can be regarded as a special form of LTL_f traces with the following characteristics:

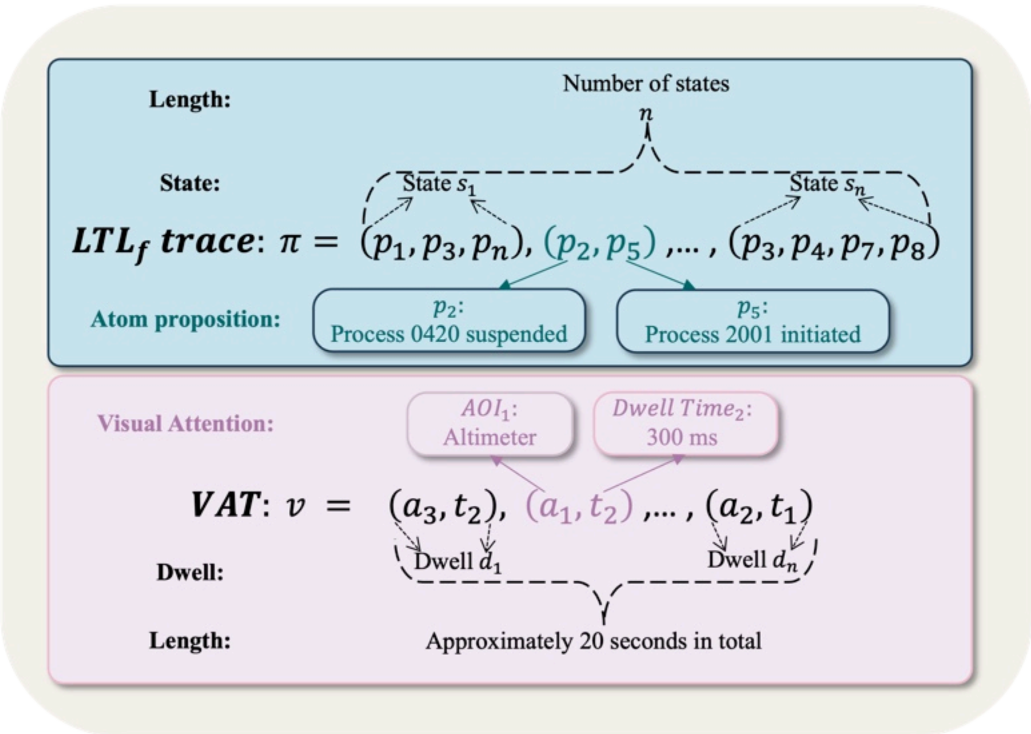

Figure 4.2: Composition of LTL_f traces and VAT

- **Basic element**: The basic elements of LTL_f traces are atom propositions

$p_i \in \mathcal{P}$ which represent fundamental facts of a system, such as "*Process 0420 suspended*" and "*Process 2001 initiated*". The number of such atomic propositions in a system can reach thousands, depending on the settings. In contrast, VAT comprises two types of basic elements: dwelled AOI $a_i \in \mathcal{A}$ and dwell time $t_i \in \mathcal{T}$. Given the instruments considered in the cockpits, the number of AOIs is likely to range from ten to twenty. Dwell time, initially a continuous factor, is processed into five binned levels in VALIO.

- **States and dwells**: Each state $s_i \in \mathcal{S}$ in LTL_f traces includes several atomic propositions (p_i) depending on the number of events (which can be zero) occurring at the selected moment. Conversely, each dwell $d_i \in \mathcal{D}$ in VAT strictly contains only two basic elements: dwelled AOI (a_i) and dwell time (t_i). The states s_i in LTL_f traces typically represent a uniform time length in reality, based on the predefined system settings. The dwells d_i in VAT usually signify varying time lengths in reality, depending on the actual dwell time.

- **Length of trace**: The duration of LTL_f traces can span from seconds to hours, contingent upon the task at hand for analyzing system behavior. To ascertain pilots' OOTL status through visual attention analysis, VAT duration is set within a range of approximately several seconds (e.g., 20 seconds) in VALIO, facilitating the assessment of pilot behaviour and status within a specific timeframe.

Following labeling the training VATs, VALIO trains a Graph Neural Network (GNN) model to distinguish between positive and negative traces, based on the methods proposed by Luo et al. [91]. Each VAT (v) is converted into a directed graph $G_v = (N_v, E_v)$ for application of GNNs. Here N_v is the set of nodes in the graph, with each node n_i corresponding to a dwell d_i in v. Each pair of adjacent dwells (d_i, d_{i+1}) in VAT corresponds to a pair $< n_i, n_{i+1} >$ in E_v, which is the set of edges in the graph. For each node $n_i \in N_v$, an associated feature vector with a fixed-length encodes information about the propositions and sub-formulas of non-atomic propositions in the VAT. The trace graphs (G_v) representing the VATs with the feature vectors are used as input to train the GNN model. The GNN model is trained using a binary cross-entropy loss function to minimize the classification error.

Once the GNN model is trained, the LTL_f formula is extracted by interpreting the parameters of the learned GNN classifier. This interpretation process involves analysing the weights and biases of the GNN model to identify the sub-formulas that contribute most to the classification decision. These sub-formulas are then combined to form the final LTL_f formula.

4.2.3 OOTL Identification

After obtaining the final LTL_f formula from the training VATs, the OOTL status can be identified by verifying the given VATs. In addition to the identification of OOTL, a highlight the VALIO framework is the extra insights into the OOTL phenomenon brought by analyzing the verification results.

By training the GNN model with different training data from the collected dataset, multiple different LTL_f formulas might be obtained. In addition to the explanations provided by the human-readable formulas themselves, it is also possible to acquire further understanding of OOTL from the testing results. More specifically, different LTL_f formulas might achieve similar classification accuracies, suggesting the different visual behaviours represented by those LTL_f formulas are both correlated to the OOTL status. But with different precision and recall, the interpretation of such behaviours might be different. Consequently, it is possible to obtain further insights from the formulas and therefore better understand the characteristics of the OOTL phenomenon.

4.3 Case study

The section evaluated the VALIO framework through a flight simulation experiment conducted at Hong Kong Polytechnic University, involving 26 participants (17 males, 9 females, aged 22-32). This section details the methodology, data collection, and performance evaluation of VALIO, and compares the efficacy of using various VAT lengths. It also contrasts the VALIO against other widely used methodologies, such as Random Forest, XGBoost, and Multilayer Perceptron Neural Networks.

4.3.1 Experiment design and data collection

The experimental framework utilized a Cessna 172 simulator, coupled with Microsoft Flight Simulator [153] to create a realistic aerial environment. A Tobii Pro Glasses 3 [154], with 16 illuminators and 4 eye cameras was utilized to track the participants' eye movements. The eye tracker also holds a scene camera in the front to record the field of view, so the eye movements can be mapped as gaze behaviors in the scene. Meanwhile, a 27-inch monitor on a desktop computer facilitated data collection, as depicted in Figure 4.3. Among the 26 participants, four of them had prior experience (less than 5 hours) with the flight simulator, while the remaining participants were novices. All of the participants were right-handed and had normal or corrected-to-normal vision. Ethical approval for the study was granted by the PolyU Institutional Review Board (Reference number: HSEARS20211117002), and written informed consent was obtained from all subjects before the experiment.

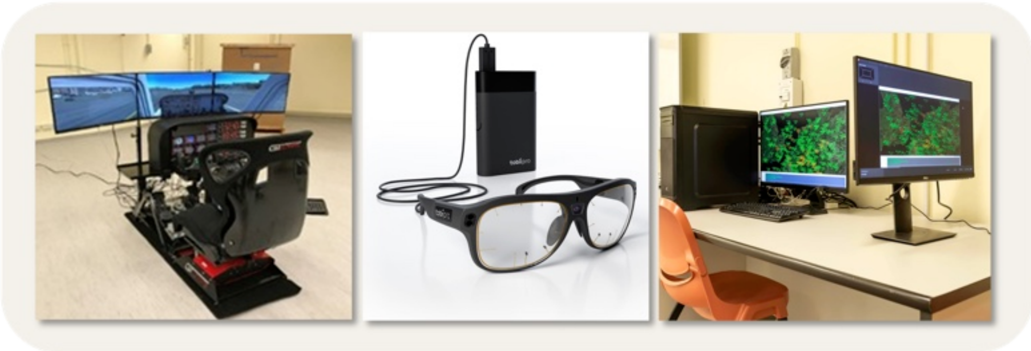

Figure 4.3: Experiment apparatus

The experiment procedure began with an introductory session, followed by a practice session until participants comfortably mastered the simulator's flight tasks. To ensure the participants' capabilities in managing flight control, the experiment adopted a simplified manual operation requirement for the participants under the visual flight rules (VFR) [21]: they only need to conduct a Straight-and-Level flight [112] by maintaining pitch, roll, and yaw angles using joysticks and throttle through instrument panels and external views. The other tasks like navigation and fuel monitoring were automated and no effort was required from the participants.

Participants continued practising until they could independently perform a complete flight procedure, including take-off, a 25-minute cruise, and landing. The formal experiment follows a one-way within-subject experimental design, requiring each participant to complete two identical flight sessions. Each session commenced with a pre-saved cruising phase at 2,800 ft AMSL record under autopilot mode, lasting about 42 minutes while maintaining the altitude. The meteorology condition was set as the "Clear Skies" in the Microsoft Flight Simulator, with a wind of 18 gusts/minute, average speed of 1 knot, and no cloud below 10,000 ft AMSL. In one session, continuous autopilot engagement was maintained to induce Out-Of-The-Loop (OOTL) status in participants, with the final 120 seconds marked as OOTL. This approach aligns with prior research indicating that over 20 minutes of automation exposure reliably triggers the OOTL phenomenon [49, 116, 155]. The alternate session involved manual aircraft control for two minutes at two separate intervals, with data during these intervals labeled as In-The-Loop (ITL). Participants were unaware of the timing and presence of manual control requirements before the flight. Figure 4.4 provides a visual demonstration of the two flight sessions. Prior to the formal experiment, the eye tracker was calibrated for each participant by directing them to look at the calibration dot on a calibration card, following the process of the Tobii Field Guide [156]. After the first simulation flight, a break of 10 to 20 minutes based on the participants' need was given to mitigate fatigue effects on the subsequent trial.

4.3.2 Learn LTL_f formulae with Visual Attention Traces

Fourty-two records from twenty-one participants were selected and processed by excluding data from 5 participants (10 records) due to the poor eye-tracking rate being below 85% to ensure data quality [157–159]. We tested four methods for

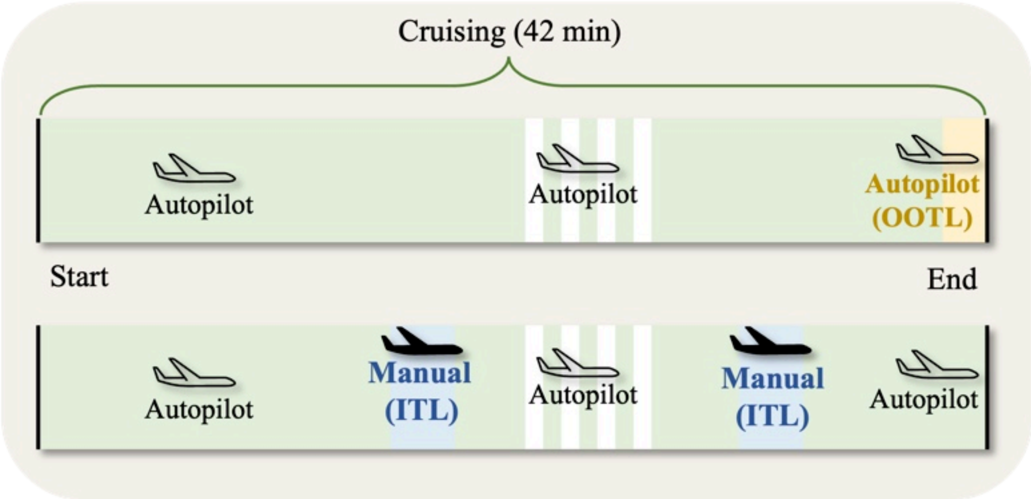

Figure 4.4: Flight tasks

dividing the dwell times: quartiles, logarithm, Gaussian mixture model (GMM), and K-Means clustering. Meanwhile, we identified six Areas of Interest (AOIs) crucial for monitoring aircraft status and executing flight control tasks. Then the Visual Attention Traces (VATs) were complied for the LTL_f processing.

4.3.2.1 Areas of Interests in the cockpit

The Cessna simulator provides instruments on the panel and the simulated view of the flight. Figure 4.5 shows a screenshot when processing an eye tracking record in the *Tobii Pro Lab* software. The left part on the screen presents the scene video recorded by the eye tracker, with a gaze moving from the external view to the panel. The right part on the screen is a photo imported into the software for mapping the gaze behaviors. The AOIs are defined based on this statical picture, and hence the eye movements recorded by the eye tracker can be mapped on the picture as dwells into AOIs. Specifically, this study defined six AOIs as depicted in the upper part of Figure 4.7. These AOIs encompass key instruments and visual cues within the flight simulator. Specifically, they include the attitude indicator (ATT), altimeter (ALT), airspeed indicator (SPD), vertical speed indicator (VSPD), the aircraft's nose (Nose), and the view outside the window (OTW). The ATT, ALT, SPD, and VSPD provide digital information about the aircraft's status, while the Nose and OTW offer a visual representation of the aircraft's orientation relative to the horizon.

Figure 4.5: Screenshot of Tobii Pro Lab

4.3.2.2 Divide dwell times into different levels

To encode the dwell time into atomic propositions that are compatible with LTL_f methods as well as trying to retain the reflection of how the participants allocate their attention, we tried to bin the dwell times into 5 levels using quartiles, logarithm, GMM, and K-Means clustering methods. Specifically, though most eye trackers can capture gaze movements at a frequency of over 50 Hz, viewing a stimulus for less than 100ms is typically not counted as an effective fixation or dwell

[160]. Therefore, all the dwells shorter than 100ms are firstly classified to the T_1 level. Then the rest dwell time data was binned into four dwell levels ($\mathcal{T}$) using the methods below:

- *Quartiles*: Define the bin edges using the 25th percentile (241ms), 50th percentile (491ms), and 75th percentile (1052ms) of the dwell time distribution.
- *Logarithm(Ln)*: Define the bin edges using logarithm to the base of the mathematical constant e. Considering $e^{4.6} \approx$ 100ms (T_1 level), the rest bin edges are $e^{5.6}$ ($\approx$ 270ms), $e^{6.6}$ ($\approx$ 735ms), and $e^{7.6}$ ($\approx$ 1998ms).
- *GMM*: Define the bin edges by applying the Gaussian mixture model to cluster the dwell times. Specifically, the edges are determined as 525ms, 1728ms, and 5832ms.
- *K-Means*:Define the bin edges using the K-Means clustering method. Consequently, 1734ms, 6011ms, and 17531ms are used as bin edges.

Figure 4.6 provides a visualized illustration for the distribution of the collected dwell times in this case study and the divisions using different methods. (There are some single dwells longer than 20,000 ms that are omitted in the figure for layout, with the longest one of 70,789 ms)

4.3.2.3 Encoding VATs

A dwell $d_i(a_i, t_i) \in \mathcal{D}$ in this study can be defined by the interaction bewteen the 6 AOIs and 5 binned levels, yielding 30 possible combinations. VATs were constructed based on these combinations, following the sequence of events as shown

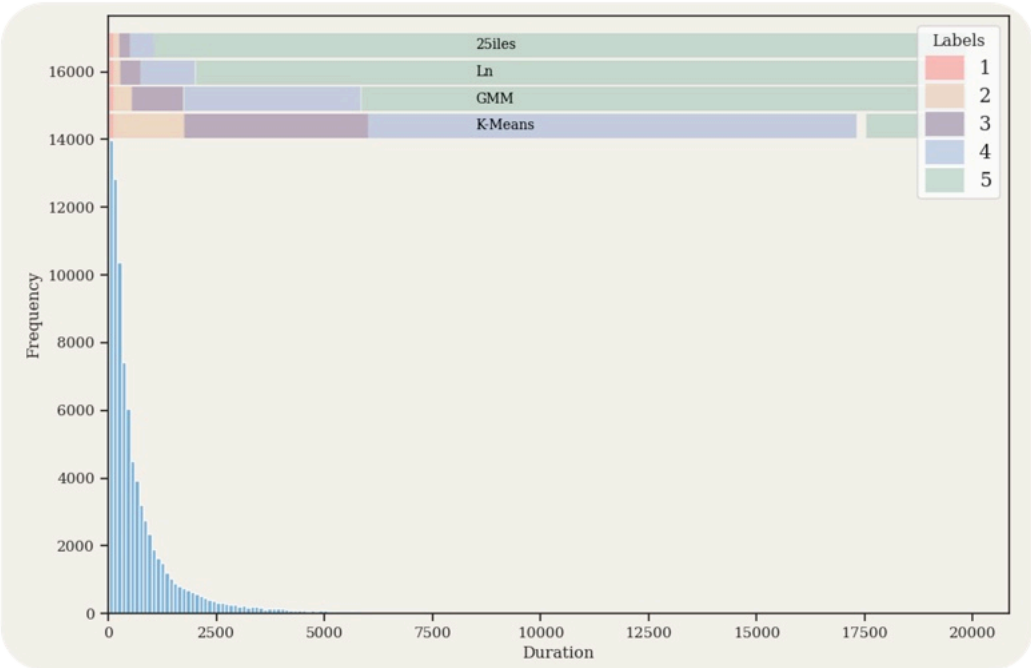

Figure 4.6: Division of dwell times using different methods

in Figure 4.7. Based on the previously established labels, the VATs were categorized into OOTL status and ITL status. Based on the work of Luo et al. [91], an aggregate-combine Graph Neural Network (AC-GNN) model was constructed and trained using these VATs (*epoch = 60*). The parameters of the trained GNN model were then interpreted to obtain the LTL_f formulas for discerning OOTL status.

4.3.3 Results and comparisons

This case study tested the VALIO method using time windows from 10 seconds to 75 seconds (increase 5 seconds every time) to verify its effectiveness across different time windows (VAT length). All four dwell time division methods with different time windows were tested using a 10-fold cross-validation method. Pre-

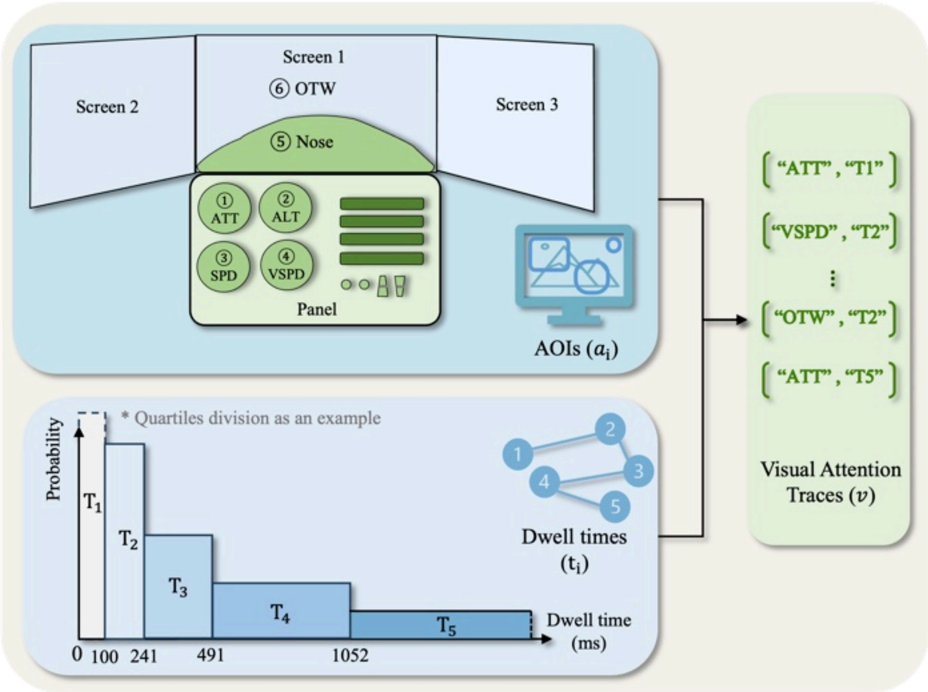

Figure 4.7: Temporal-spatial information in the VATs

cision, recall, and F1 scores were adopted as performance metrics to compare the outcomes. Precision denotes the accuracy of positive predictions, while recall indicates the rate at which actual positives are correctly identified. The F1 score harmonizes precision and recall into a single metric, measuring a model's balanced accuracy in positive prediction and identification.

4.3.3.1 Identified LTL_f formulas

The VALIO framework was tested with 56 combinations (4 division methods $\times$ 14 time windows) using 10-fold cross-validation. In the GNN training and parameter interpretation phase of each round, the VALIO framework generates one LTL_f formula and uses this normal expression to classify the testing data for OOTL identification. These LTL_f formulas serve as distinct classification models in the single round with their individual performances on the testing dataset. Specifically, three LTL_f formulas were recognized in these 560 rounds: "$G\neg\ att$", "$G\neg\ alt$", and "$G\neg\ vsp$". The distribution and performance of these three formulas on different time windows are presented in Figure 4.8. The figure demonstrates that the occurrence of these formulas is significantly related to the length of the time window. It is noticeable that when the overall performance (i.e., F1 score) of the previous formula declines, the other formula emerges with a better performance. Meanwhile, these formulas were recognized across different dwell time division methods with no significant difference between division methods. The interpretations and overall performance of the formulas are presented as follows.

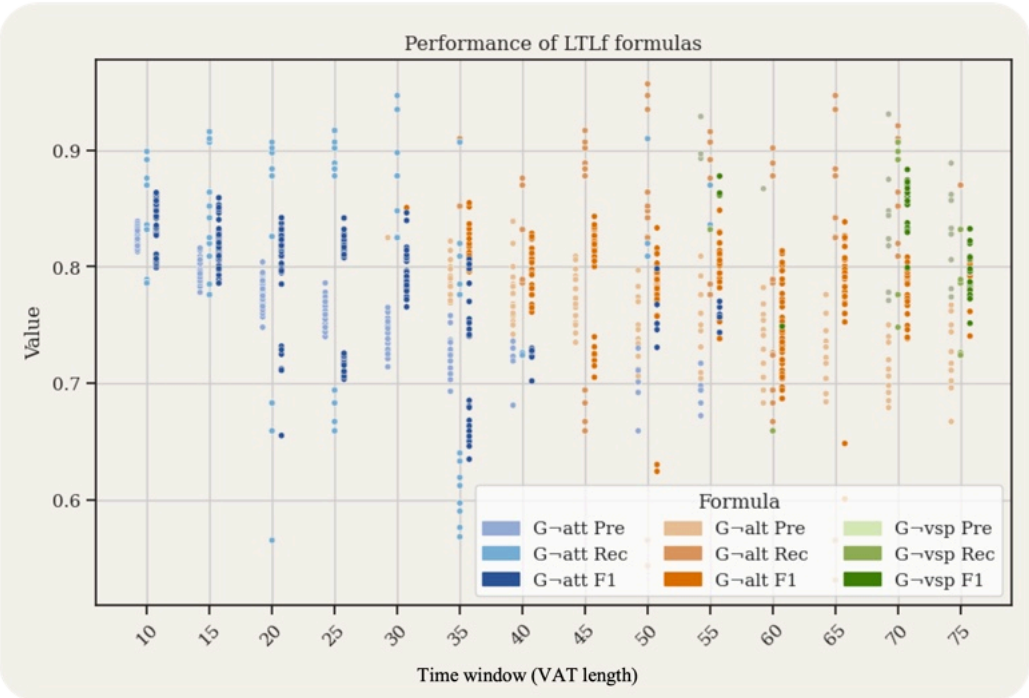

Figure 4.8: Distribution and performance of LTL_f formulas

- "$G\neg\ att$": This formula signifies that the pilot's attention was never directed towards the attitude indicator (ATT) throughout the whole interval. It dominates the time windows from 10 seconds to 30 seconds, and it is also recognized with decreased performance for the time windows between 35 seconds to 55 seconds. It was recognized 279 times in total with an averaged Precision of *0.75*, Recall of *0.83*, and F1 score of *0.79*.

- "$G\neg\, alt$": This formula indicates that the pilot did not focus on the altimeter (ALT) at any point during the whole interval. It was recognized from 35 seconds to 70 seconds, with a total occurrence of 237 times. The averaged Precision is *0.77*, Recall is *0.82*, and F1 score is *0.79* across all time windows and dwell time division methods.
- "$G\neg\, vsp$": Similar to the previous two, this formula implies that the pilot didn't look at the vertical speed indicator (VSP) during the whole interval. It emerged when the time window is longer than 55 seconds. It obtained Precision of *0.84*, Recall of *0.81*, and F1 score of *0.82* with a total frequency of 44 times.

4.3.3.2 Sensitivity of time windows and division methods

The effectiveness and stability of the VALIO across different time windows were verified using different dwell time division method. As the 10-fold cross-validation was adopted to test each combination of the time window and dwell time division method, different LTL_f formulas might be recognized in these 10 rounds. For example, "$G\neg\, alt$" was recognized 4 times and "$G\neg\, vsp$" was recognized 6 times for the 75-second time window when using quartiles to bin the dwell times. And the performance metrics were averaged using the individual performances of these ten times. The result indicates that the performance is stable with no significant differences between different time windows for the four tested dwell time division methods, as shown in Figure 4.9. Specifically, the lowest F1 score (*0.769*) was identified with the Precision of *0.819* and Recall of *0.725* when using the K-Means clustering method and 75-second time window. The highest F1 score (*0.815*) was obtained with Precision of *0.777* and Recall of *0.857* when using the Logarithm method and 45-second time window.

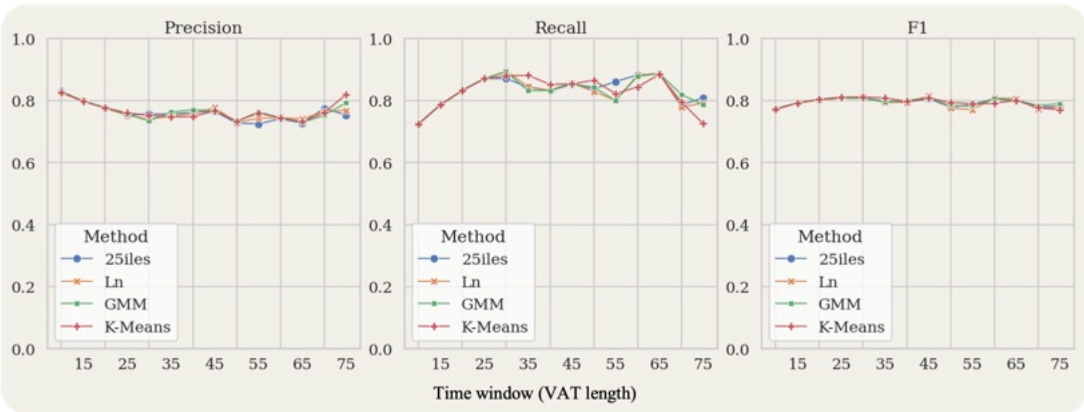

Figure 4.9: VALIO performance using different dwell time division methods

4.3.3.3 Comparison with other methods

We compared the performance of the VALIO framework (using the Logarithm division method of dwell time) against seven well-established predictive models: Hidden Markov model (HMM), Linear Discriminant Analysis (LDA), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Gradient Boosting Decision Tree (GBDT), Light Gradient Boosting Machine (LGBM),and Multilayer Perception (MLP), using identical data of the 14 time windows. Different from using the binned dwell times in VALIO, these models were trained directly with the dwell duration in milliseconds. All these models were constructed using *Python 3.9*, with their major hyperparameters illustrated in Table 4.1.

The performance of the compared methods is shown in Figure 4.10. The result indicates that the RF, XGBoost, GBDT, LGBM, and MLP methods exceeded the performance of VALIO in Recall and F1 score when the time window was larger than 30 seconds. The best F1 score (*0.849*) was achieved by the GBDT method when using the time window of 55 seconds, with the Precision of *0.760* and Recall of *0.962*. However, the VALIO outperformed the other models in Recall rate and F1 score when the time window was smaller. It kept a stable performance with an average F1 score of *0.793* across all the tested time windows.

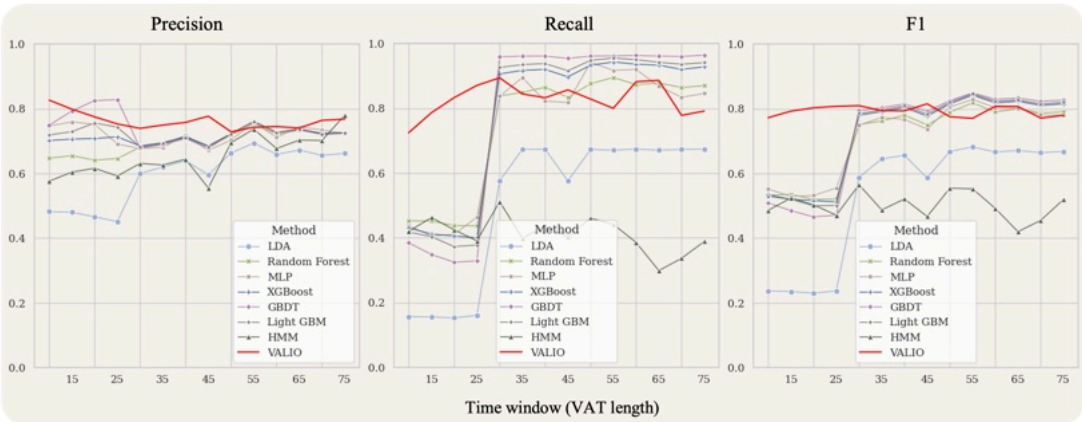

Figure 4.10: Performance of VALIO compared with other methods

4.3.4 Discussion

This case study demonstrates the efficacy of the VALIO framework in encoding pilots' visual attention into VATs and identifying OOTL status. By comparing the performance across various VAT lengths, the VALIO method demonstrated a stable performance with F1 scores between 0.77 and 0.82. Meanwhile, the comparison

Table 4.1: Machine learning models and the major hyper-parameters

Models and Library	**Hyperparameter**	**Function**	**Value**
HMM hmmlearn (v-0.3.2)	n_components	Number of states	2
	covariance_type	Each state uses diagonal covariance matrix	”diag”
	n_iter	Maximum number of iterations	100
LDA scikit-learn (v-1.3.2) .discriminant_analysis	solver	Singular value decomposition	”svd”
	shrinkage	No shrinkage	None
RF scikit-learn (v-1.3.2) .ensemble	n_estimators	Number of trees in forest	100
	min_samples_split	Minimum samples to split an internal node	2
GBDT scikit-learn (v-1.3.2) .ensemble	loss	Binomial and multinomial deviance	”log_loss”
	learning rate	Shrinkage of each tree’s contribution	0.1
	n_estimators	The number of boosting stages	100
MLP scikit-learn (v-1.3.2) .neural_network	hidden_layer_sizes	Number of neurons in the ith hidden layer	(100,)
	activation	Rectified linear unit function	”relu”
XGBoost xgboost (v-2.0.3)	tree_method	Method for constructing the trees	”hist”
	early_stopping_rounds	Enables early stopping	2
LGBM lightgbm (v-4.3.0)	learning_rate	Shrinkage of each tree’s contribution	0.1
	n_estimators	The number of boosting stages	100

with other classification models also indicates that the VALIO framework outperforms when using shorter time windows with higher recall rates and F1 scores than other methods. This highlights its capability to identify pilots' OOTL status in time and efficiently.

4.3.4.1 VAT lengths and dwell time divisions

The VALIO method obtained three LTL_f formulas and they contributed to a stable performance of VALIO across various time windows. These formulas provide not only OOTL identification capability, but also provide insights by its explainability.

The formula "$G\neg\ att$" emphasizes the critical role of the attitude indicator (ATT) in ensuring pilot engagement. Considering the great importance of aircraft attitude adjustments (pitch, roll, and yaw angles), the pilots need to regularly check it with a high frequency. Therefore, neglecting this indicator is identified as a strong predictor of OOTL status. However, a slight decline in its occurrence and prediction performance was observed when the time window increased, which might be explained by the increasing possibility of the dwell towards attitude indicator, considering its necessity in long-term flight monitoring activities.

The formula "$G\neg\ alt$" underscores the importance of routinely checking the altimeter. The change in altitude is less extensive than the attitude, so the need to check the altimeter is less than checking the attitude indicator. Therefore, this formula was recognized as the OOTL predictor at a medium frequency with time windows between 35 to 70 seconds.

Furthermore, the formula "$G\neg\ vsp$" was recognized a few times when the time window became longer than 50 seconds. Comparing to the attitude indicator and the altimeter, the vertical speed is less checked during the cruising phase. Therefore, the dwells toward the speed indicator suggest more active scanning activities and can signify the ITL status. This is in line with research by Di Stasi et al. (2016) and Diaz-Piedra et al. (2016), which links reduced saccade rates and velocities to increased fatigue and decreased vigilance [67, 68].

These three formulas were recognized with the performance decline of the previous formula as shown in Figure 4.8, resulting in a stable overall performance of the VALIO methods across time windows.

4.3.4.2 VALIO and other classification methods

As shown in Figure 4.10, some other tested models (i.e., GBDT) obtained better Recall rates and F1 scores when using the longer time windows. However, the VALIO method outperformed in three perspectives: better performance with shorter time windows; more stable performance across different time windows; and explainability by human-readable LTL_f formulas. This demonstrates that the proposed VALIO method is expected to identify the OOTL status of the pilot with better timeliness and explainability, and therefore enables more in-time interventions to prevent the negative effects.

Meanwhile, a significant improvement in the recall rates and the F1 scores of other models was observed in Figure 4.10 when the time window was longer than 30 seconds. Another observation from Figure 4.8 indicates that more LTL_f formulas started to arise when the time window was longer than 30 seconds. A possible explanation can be made that the visual monitoring behaviors of these student pilots have a periodicity. The VALIO method can handle the characteristics with either more or fewer periodical cycles by discovering different formulas, while the other methods might need the eye-tracking data of more periodical cycles to capture the characteristics. However, this is a hypothesis that needs further investigation by more empirical studies in the future.

In summary, the results demonstrated that the VALIO framework effectively translates eye-tracking data into VATs and employs LTL_f techniques to generate human-readable formulas for OOTL detection. It fills the critical gap in explainability associated with detecting OOTL status and provides insights into understanding the OOTL status. This methodology surpasses other state-of-the-art approaches with shorter time windows and more stable performance, showcasing its ability to detect OOTL status with better timeliness.

4.4 Concluding remarks

The phenomenon of Out-Of-The-Loop (OOTL) is a prevalent concern in aviation, often induced by high levels of automation. It significantly impairs pilot performance and aviation safety. However, a notable challenge persists in the explainable characterization and identification of pilots' OOTL status. Addressing this, our study introduced the *Visual Attention* LTL_f *for Identifying OOTL (VALIO)* framework, utilizing eye-tracking data to discern pilots' OOTL status with enhanced explainability.

This research makes three significant contributions to addressing the gap of explainability. First, it introduces an innovative method for encoding pilots' eye-tracking data into structured Visual Attention Traces (VATs), which capture the temporal and spatial dynamics of visual attention. These VATs effectively represent pilots' information-gathering behaviors, providing a solid foundation for explainability in the analysis. Second, the study utilizes LTL_f methods to analyze these VATs, successfully identifying pilots' OOTL status with explainable results by generating human-readable formulas. Human-readable formulas could reveal what scanning behaviors have been identified as OOTL, thus providing direct instructions on how the pilots should adjust their scanning behaviors. In comparison with other methods, this methodology significantly outperforms when using shorter time windows and hence provides better timeliness. Third, by conducting a focused case study, the VALIO framework generated three LTL_f formulas for the given flight task. These formulas offer valuable insights into the characteristics of OOTL status, deepening our understanding and directing future research efforts.

However, our study has limitations. The proficiency level of student pilots involved and the constrained scope of laboratory-collected data limits the findings from being directly adopted to the practical conditions. Furthermore, the binary classification of OOTL and ITL statuses in this study simplifies a more nuanced reality. Although having these two limitations, the results demonstrated that the proposed method is able to capture the characteristics of the pilots' gaze movement for discerning different statuses. Future research aims to encompass a broader spectrum of pilot expertise, incorporating licensed pilots and more realistic scenarios. Moreover, the dwell time division method, as well as the training and interpretation methods for LTL_f formulas will be further developed to enable more comprehensive expressions and better performance. This will enable a more extensive exploration of OOTL's varying degrees and enhance the robustness of our findings.

Chapter 5

Study 3: A context-aware ITL support approach

This chapter presents a work to address the challenge in interactions with the ITL pilots. The proposed method tokenizes the eye-tracking data into Visual Attention Matrices (VAMs) and integrates with a Large Language Model (LLM) to identify and respond to troubleshooting activities of ITL pilots. This chapter is organized as follows: Section 5.1 introduces the background of this work and identifies the research gaps. Section 5.2 details the proposed VAM framework and its integration with LLMs. Section 5.3 presents a case study validating the proposed approach. Section 5.4 discusses the implications and limitations of this work. Finally, Section 5.5 makes concluding remarks for this work.

5.1 Background

As discussed in Section 2.2.2.3, the dynamic and comprehensive needs of the ITL pilots poses a research challenge in the interaction of automation support. This work selects the troubleshooting behaviours of the pilots as a scenario to address this challenge. The troubleshooting activities are complex and blended with normal monitoring behaviours. It may encounter cross-checking various instruments in the cockpit, which is similar to the normal monitoring behaviours. It is difficult to distinguish these two statuses, especially when trying to identify where the incident occurs. Though researchers have developed many methods to assess operator status and predict performance using eye-tracking data [33], these evaluations typically represent a single-dimension evaluation from low to high (e.g., situation awareness, fatigue). These methods can hardly distinguish troubleshooting from normal monitoring and identify where the incident occurs, therefore hindering the user-friendly interaction of automation support. To address this, we explore analyzing the eye-tracking data with the trending Large Language Model (LLM) GPT-4, which has demonstrated remarkable abilities in handling complex context and offering natural interactions [84, 85].

However, a gap lies in effectively processing eye-tracking data with the LLM due to the lack of semantic information in eye-tracking data. Although LLMs have demonstrated the ability to identify deviations from patterns in time series logs [161, 162], their application for processing eye-tracking data remains underexplored. The abilities of LLMs are built on processing natural languages, which contain ample context and semantic information [163]. Eye-tracking technologies produce high-frequency data, with sampling rates reaching up to 2000 Hz.

It leads to vast datasets, while no semantic information [164, 165]. This lack of semantic information in eye-tracking data challenges the reasoning capacities of LLMs [166]. Meanwhile, training the LLM with all empirical data and inputting the current eye-tracking data segment, representing the ongoing visual attention, can place a significant burden on the LLMs' data processing and reasoning ability. To address the issues, we developed a visual attention matrix (VAM), which compiles the temporal-spatial information from the eye-tracking data into a 9x9 matrix, providing concise, tokenized input for GPT-4 model.

In this work, we contribute to the HCI in aviation from two perspectives. First, we use LLM to process eye-tracking data, enabling a more comprehensive and context-aware estimation of pilots' troubleshooting. This approach enhances HCI by enabling discrimination of pilots' troubleshooting activities from blended status with normal monitoring behaviors. Second, we developed the Visual Attention Matrix (VAM) to summarize semantic information in eye-tracking data, enabling its tokenization for input to LLMs. VAM represents a novel approach for summarizing temporal-spatial information in eye-tracking data and provides a reference for using LLMs with non-semantic psychological datasets. In Section 5.4, we verified the ability of our method to identify and respond to the pilots' troubleshooting in their ITL status with a case study. Together, these contributions offer insights into enhancing HCI by enabling a more user-friendly interaction from identifying the ITL pilots' needs for support and responding to it. This enhancement aims to foster a more comprehensive and robust closed-loop automation support mechanism.

5.2 Visual Attention Matrix (VAM)

To identify the pilots' information needs using LLM and eye-tracking data, we developed the Visual Attention Matrix (VAM) to tokenize the eye-tracking data and designed prompts to integrate with the LLM. This section details how to compile the eye-tracking data into VAMs and integrate it with LLM.

5.2.1 Compiliing eye-tracking data into VAM

We introduce the VAM as a structured representation to analyze the pilot's effort in information acquisition. It summarises the fixation and saccade activities into a matrix, reflecting not only the fixation durations of each stimulus but also the transitions between the stimuli. Figure 5.1 provides a graphical illustration of constructing VAM from eye-tracking data.

Firstly, the key instruments in the cockpit are defined as Areas of Interest (AOIs), with the total number denoted as N. Correspondingly, we define a matrix $\mathcal{M}_{N\times N}$, where both the nth row and n^{th} column represent the same AOI ($1 \leq n \leq N$).

Secondly, each fixation event is extracted from the eye-tracking data, described as (i, p, q, x), where: $i \in I$ is the index of the fixation event in the sequence, $p \leq N$ indicates the previous fixated AOI, $q \leq N$ indicates the current fixated AOI, and x represents the fixation duration (in milliseconds) on the current AOI.

Next, the fixation events are then encoded into the matrix $\mathcal{M}_{N\times N}$ in an additive manner. Specifically, each element $\mathcal{M}_{p,q} \in \mathcal{M}_{N\times N}$ is given by:

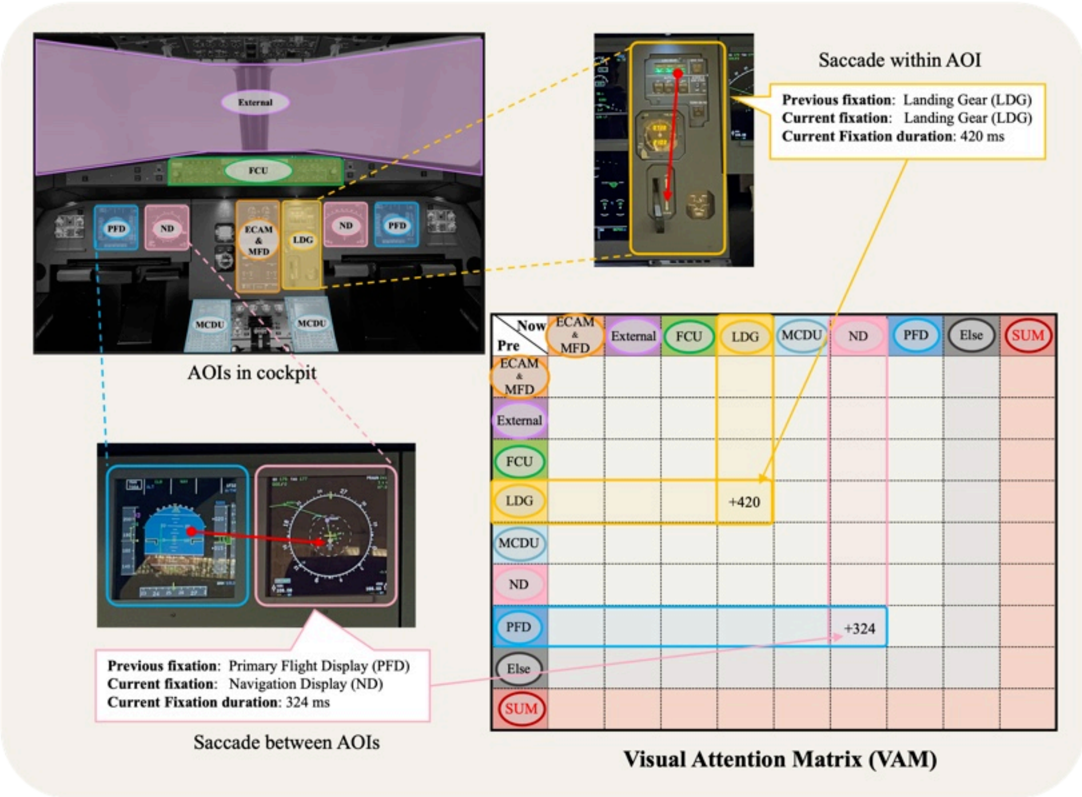

Figure 5.1: Construct Visual Attention Matrix (VAM) from eye-tracking data

$$\mathcal{M}_{p,q} = \sum_{j \in I,(j,p,q,x)} x_j \tag{5.1}$$

where x_j represents the fixation duration for each event.

Finally, the matrix $\mathcal{M}_{N\times N}$ is extended to $\mathcal{M}^{+}{}_{(N+1)\times(N+1)}$ by adding a sum row and a sum column at the end. The sum row (row $N+1$) and the sum column (column $N+1$) represent the sum of the previous rows and columns, respectively. Specifically, each element in the sum row is given by equation 5.2 and each element in the sum column is given by equation 5.3:

$$\mathcal{M}^{+}{}_{N+1,q} = \sum_{p=1}^{N} \mathcal{M}_{p,q}, \quad 1 \leq q \leq N \tag{5.2}$$

$$\mathcal{M}^{+}{}_{p,N+1} = \sum_{q=1}^{N} \mathcal{M}_{p,q}, \quad 1 \leq p \leq N \tag{5.3}$$

The element at the bottom right corner of the extended matrix, $\mathcal{M}^{+}{}_{N+1,N+1}$, represents the total sum of all elements in the original matrix:

$$\mathcal{M}^{+}{}_{N+1,N+1} = \sum_{p=1}^{N} \sum_{q=1}^{N} \mathcal{M}_{p,q} \tag{5.4}$$

Through this method, the VAM enhances the representation of visual attention by incorporating additional contextual information. The sum row represents the total fixation duration for each AOI, offering a fundamental metric of visual attention distribution. Similarly, the sum column indicates the total fixation duration directed toward subsequent AOIs after viewing the current AOI, providing a novel metric for assessing the interaction between AOIs. Specifically, each element $(\mathcal{M}_{p,q})$ in the matrix represents the weight of transitions between AOIs, thereby offering more comprehensive insights into the pilot's information acquisition effort beyond merely calculating total fixation durations on individual AOIs.

5.2.2 Detecting pilots' information needs with LLM and VAM

The VAM introduced in the previous steps provides a concise representation for tokenizing eye-tracking data, capturing key aspects of pilots' visual attention allocation and transitions. Based on the VAM, we use the trending LLM, GPT-4, to detect and respond to the pilots' ITL confusion. The integration of VAM and

GPT-4 model is illustrated in Figure 5.2.

First, we employ three types of empirical VAMs to distinguish between pilots' normal monitoring activities and cases when they are actively acquiring specific information for troubleshooting: Normal Range VAM, Abnormal Range VAM, and P-value VAM. Each element $(\mathcal{M}^{+}{}_{p,q})$ in the Normal Range VAM and Abnormal Range VAM is a 2-tuple, (min, max), representing the minimum and maximum values derived from empirical data. The Normal Range VAM uses data from routine flight periods when no incidents occur, and the ITL pilot is engaged in standard monitoring activities. In contrast, the Abnormal Range VAM uses data from cases when abnormalities arise, and the ITL pilot is actively acquiring information for troubleshooting. Overlaps may occur between these two ranges, particularly with the lower limits, as the differences in troubleshooting are primarily reflected in the upper limits, indicating increased visual attention from the pilot. The P-value VAM is constructed by performing T-tests between the corresponding elements from normal and abnormal cases. Elements with P-values below 0.05 suggest significant differences between normal and abnormal scenarios. These elements can thus provide valuable insights for the GPT-4 model, helping it estimate whether the pilot is engaged in troubleshooting.

Second, we use the GPT-4 model to distinguish whether a pilot is conducting normal monitoring or troubleshooting based on these different VAMs. The prompt consists of the following four parts:

- *Basic instructions*: This part provides the fundamental context, assigning the AI the role of an experienced flight instructor and ergonomist. It sets the overall goal: to determine whether the pilot's gaze patterns indicate normal monitoring or troubleshooting behavior. This fundamental context setting prompts the GPT-4 model to utilize its relevant knowledge in aviation and eye-tracking disciplines.

- *Explanation of AOIs*: This part explains the specific AOIs defined in the cockpit, detailing their functions and the type of information they provide. The descriptions help clarify the relevance of each AOI for understanding the pilot's visual attention and how different cockpit instruments contribute to the information during the flight.
- *Explanation of VAMs*: This part describes how VAM is constructed and how it represents the pilot's gaze transitions between AOIs. It explains the basic elements and the sum row and column in the matrix, which aid in understanding overall gaze patterns and the level of interaction between AOIs. It also introduces the construction of the Normal Range VAM, Abnormal Range VAM, and P-value VAM, helping the GPT-4 model understand how to utilize empirical knowledge effectively.
- *Requirement*: The final part outlines the specific requirements for analyzing the pilot's behavior, including using empirical reference files and statistical data to determine normal monitoring versus troubleshooting. It requires the GPT-4 model to estimate whether the pilot is troubleshooting any specific AOI based on the given VAM and provide a response confirming the pilot's status while proposing appropriate support. This part also describes the expected format of the output from the GPT-4 model and provides an example to guide the response structure.

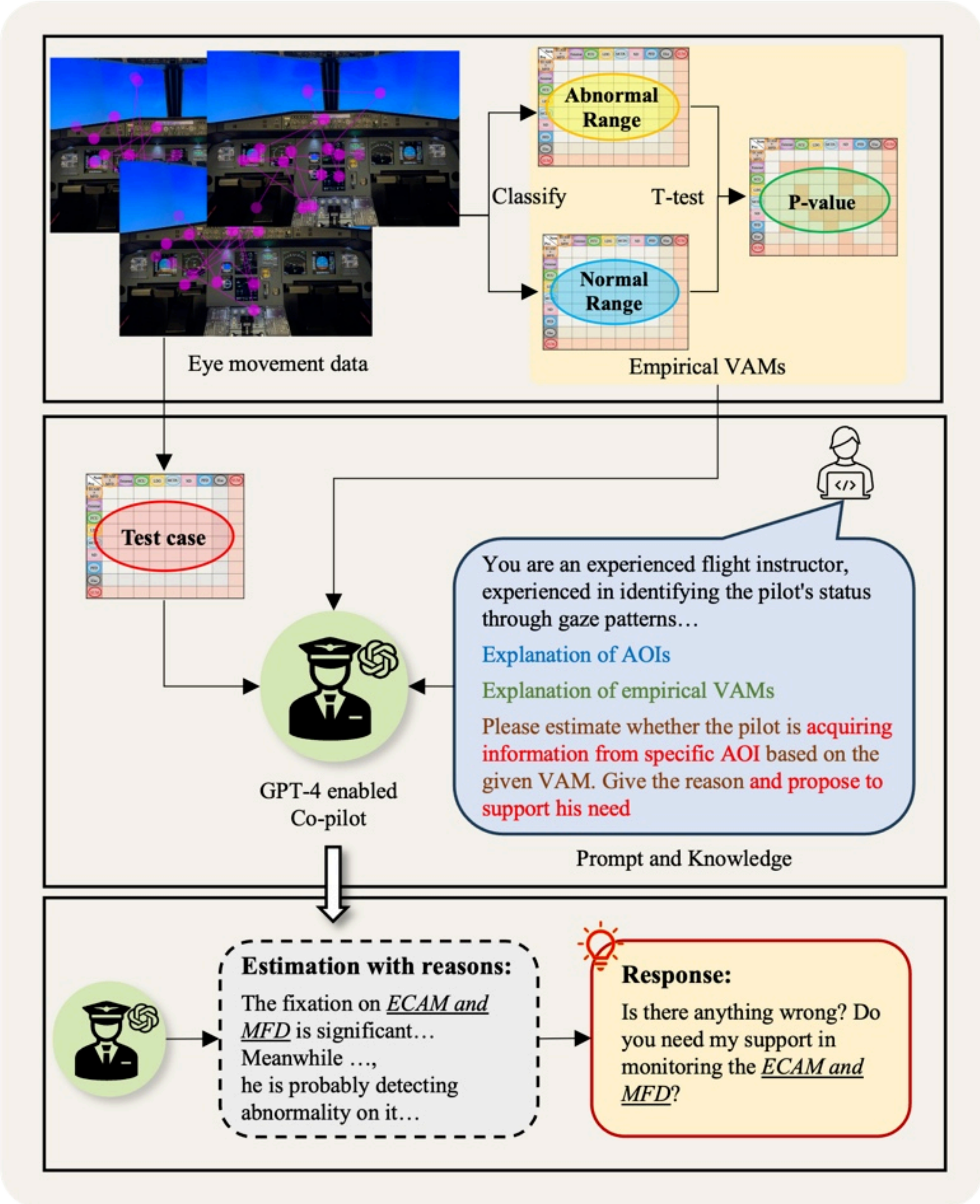

Figure 5.2: Detecting pilots' information needs with GPT-4 and VAM

This prompt helps the GPT-4 model understand the context, the pattern of input data, and the expected output. It integrates the eye-tracking data with the LLM to estimate what information the pilot requires by leveraging advanced reasoning capability, while providing a user-friendly response utilizing natural interaction capability. Meanwhile, the VAM offers a novel method to tokenize eye-tracking data as a concise input for the GPT-4 model, enhancing both processing efficiency and interpretability.

5.3 Case study

This study evaluated the proposed method through a flight simulation experiment conducted at Hong Kong Polytechnic University, involving 19 licensed pilots (all male, aged 28-55). This section outlines the methodology, data collection, and performance evaluation processes. We trimmed the VAMs (with and without sum row and column, with and without P-value reference) and compared their performance to determine the VAM format that yielded optimal results.

5.3.1 Experiment design and data collection

5.3.1.1 Apparatus and participants

The experiment utilized an Airbus A320 simulator equipped with a 180° wide-angle display to replicate the aerial environment. A desktop computer operated the *Instructor Operating Station* software, managing flight scenarios, monitoring flight status, and activating task events. Eye movement data were collected using Tobii Pro Glasses 3, featuring 16 illuminators and 4 eye cameras, with a sampling rate of 100 Hz [154]. It has a scene camera in the front to record the field of view, enabling the eye movements to be mapped as fixations to the AOIs in the scene. The setting is shown in Figure5.3. All 19 participants held valid licenses for the Airbus A320, comprising 5 Captains, 8 First Officers, and 6 Second Officers. The experiment follows ethical standards outlined in the 1975 Helsinki Declaration and received approval from the Institutional Review Board of The Hong Kong Polytechnic University (IRB Reference Number: HSEARS20211117002). Participants provided written informed consent and received HK$1,000 shopping coupons as incentives upon completing the experiment.

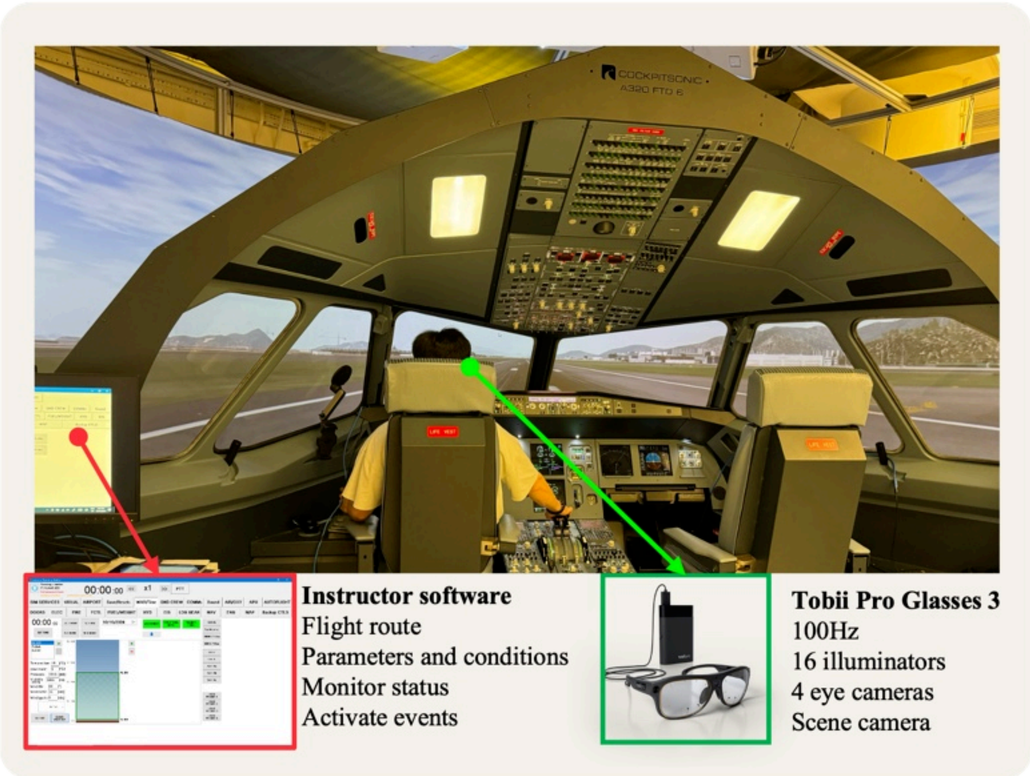

Figure 5.3: Experiment apparatus

5.3.1.2 Experiment design

The experiment involves two flight segments, simulating a round-trip, between Tokyo Narita International Airport (RJAA) and Kansai International Airport (RJBB).

The order of flight segments was randomized for each participant to ensure experimental integrity. During the experiment, a researcher acts as an Air Traffic Control Officer (ATCO) to provide instructions and responses to the pilots. Prior to the experiment, the aircraft is taxied to the takeoff position already, and the participants will get a briefing before starting. After completing the pre-flight checklist, the participant will be instructed to start the experiment from takeoff. Each segment included six phases: takeoff, climb, cruise, descent, approach, and landing, totaling approximately 50 minutes. The weather conditions are set in advance and remain the same for all the participants: wind: 080°/15 knot, runway visual range (RVR): 5,000 feet, and broken clouds between 400 and 25,000 feet.

During the flight, three distinct events were inserted as troubleshooting tasks to provide ground truth for evaluation. These tasks included an Oxygen error on the Electronic Centralized Aircraft Monitor (ECAM) and Multi-Function Display (MFD), a Traffic Collision Avoidance System (TCAS) Error on the Navigation Display (ND), and a landing gear unlocked error on the Landing Gear (LDG) panel. In default settings, these errors are accompanied by arresting audio warnings, while we muted the audios in this study to induce the participants' information acquisition visual behaviors. Participants were briefed to report any kind of incidents or abnormalities promptly during the flight. After the participant detected and reported the incident, the researchers deactivated the error from the Instructor software and instructed the participant to keep an eye on it. The unexpected lack of audio warnings and the instruction to keep an eye on it will induce the participants to pay visual attention to cross-checking, reducing reliance on audible alerts. This design raises the participant's active visual searching behaviors from specific instruments for troubleshooting. All three tasks are inserted after takeoff and before landing to avoid overload, and at least a 12-minute interval is set between every two events to avoid potential inner connections. Figure 5.4 shows the flight route from RJAA to RJBB and the three tasks. The sequence of three tasks is reversed for the flight from RJBB to RJAA.

Figure 5.4: Flight route and tasks

Apart from the designated tasks, the flight proceeded without additional events. Given the single-pilot operation (SPO) setup [110], tasks typically performed by the First Officer (FO), such as checklist reviews, were managed by the researcher. Following the first flight segment, a break of 10 to 20 minutes, depending on participant needs, was given to mitigate fatigue effects. The eye-tracker was recalibrated before the subsequent segment.

5.3.2 Data Processing

In this study, we defined 7 Areas of Interest (AOIs), including the Electronic Centralized Aircraft Monitor and Multi-Function Display (*ECAM and MFD*), External, Flight Control Unit (*FCU*), Landing Gear Panel (*LDG*), Multifunction Control Display Unit (*MCDU*), Navigation Display (*ND*), and Primary Flight Display (*PFD*). Eye movement data were mapped onto these defined AOIs using *Tobii Pro Lab* software. Additionally, a pseudo AOI called ”*Elsewhere*” was defined to account for fixations occurring outside these 7 AOIs when constructing the VAMs. The layout of the AOIs in the cockpit and the structure of the VAMs are illustrated in Figure 5.1. Detailed descriptions of the AOIs can be found in 6.2.

To optimize the format of the VAMs, we evaluated two factors: the inclusion or exclusion of sum rows and columns, and the use of an empirical P-value VAM to highlight significant elements. This evaluation led to testing four combinations: 9x9 VAMs with P-value VAM (*9-p*), 8x8 VAMs with P-value VAM (*8-p*), 9x9 VAMs without P-value VAM (*9-a*), and 8x8 VAMs without P-value VAM (*8-a*).

Thirty-one records were selected for analysis after excluding data with eye-tracking rates below 85% to ensure data quality [157, 158]. Each record included data from 3 effective tasks, resulting in a total of 93 tasks examined. We categorized the eye-tracking data into two phases: 30 seconds after the participant detected an error (Pos), indicating active information acquisition for troubleshooting; and 30 seconds before the participant detected an error (Neg), representing normal monitoring. A typical comparison of the Pos and Neg VAMs (9*9) is shown in Figure 5.5. It can be found that the task-related AOIs received more fixations after the tasks, indicating the participant’s information needs.

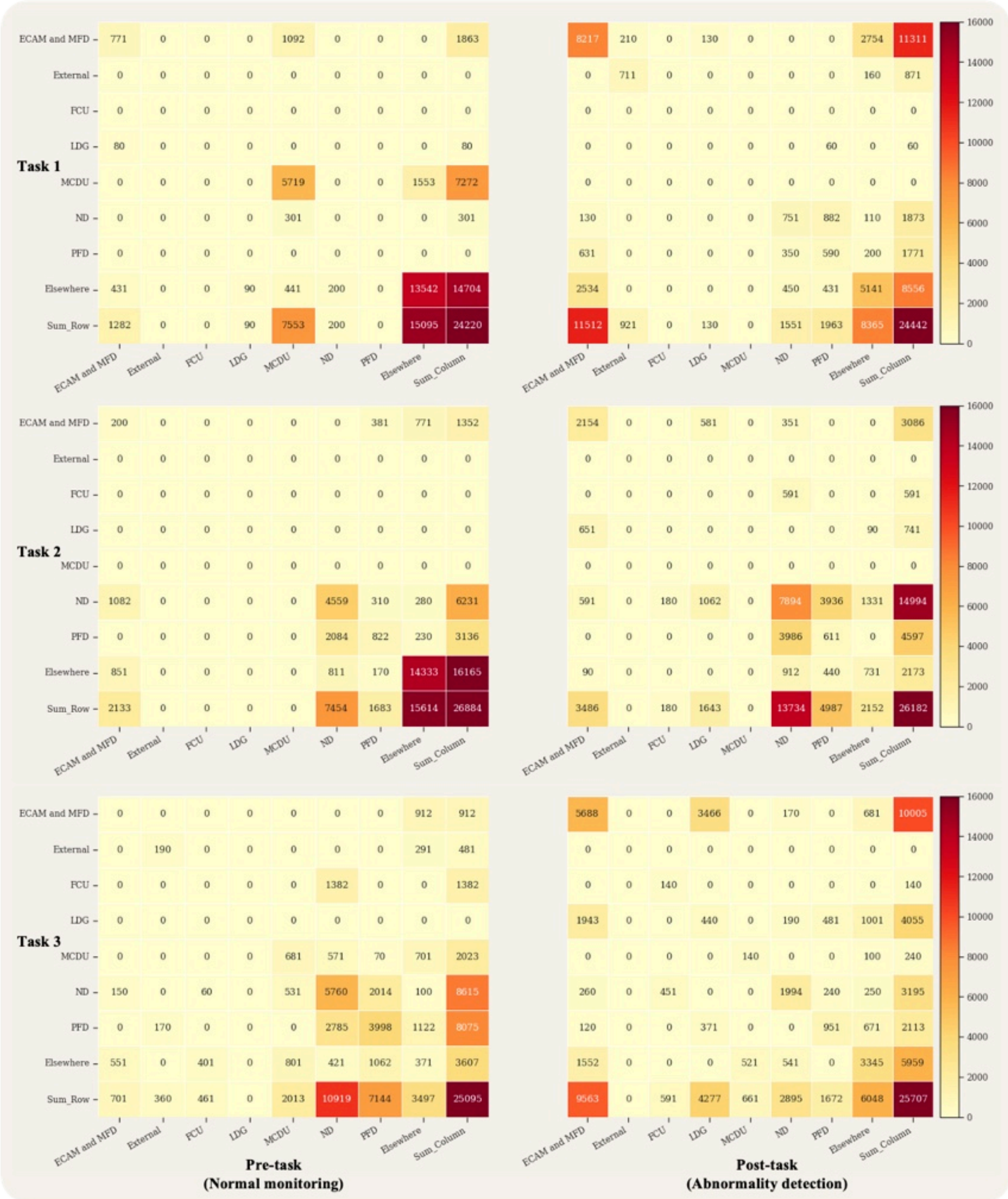

Figure 5.5: 9x9 VAMs for the three tasks (Participant 3, RJBB to RJAA)

After constructing individual VAMs for each task, we conducted customized

10-fold cross-validation using empirical VAMs to validate performance. In conventional 10-fold cross-validation, data from 9 folds are used to train the model, and the left fold data is used for testing. To avoid the heavy load of training the large model directly with massive data, we use the data from 9 folds to construct empirical VAMs (as detailed in Section 3.2) as input knowledge to the GPT-4 model. The remaining fold's VAMs were then tested to validate performance. In addition to estimating information acquisition, we tasked the GPT-4 model with providing a user-friendly response confirming participants' information needs and suggesting support options. Specifically, this study adopted the *GPT-4o* model. The prompt used in this study is detailed in Appendix 6.2, tailored to the trimmed VAMs described above.

The test was conducted from two perspectives: accuracy of the estimation and quality of the response message. Estimation performance was evaluated using macro-accuracy (Mac) and micro-accuracy (Mic) metrics. It's worth mentioning that the estimation of the GPT-4 model contains multiple possibilities, referring to the multiple AOIs. The GPT-4 model might estimate the participant was troubleshooting any of these defined AOIs. Nonetheless, the ground truth contains only four classes, including normal monitoring (Neg) before the tasks and troubleshooting behaviors after the three error detection (Pos). The three AOIs related to the tasks introduced in Section 5.3.1.2 were used as ground truth. Metrics included precision (Pre), recall (Rec), F1-score (F1), and specificity (Spec), calculated as follows:

$$Pre_{Mac} = \frac{1}{4} \sum \frac{TP_i}{TP_i + FP_i} \tag{5.5}$$

$$Pre_{Mic} = \frac{\sum TP_i}{\sum TP_i + \sum FP_i} \tag{5.6}$$

$$Rec_{Mac} = \frac{1}{4}\sum \frac{TP_i}{TP_i + FN_i} \tag{5.7}$$

$$Rec_{Mic} = \frac{\sum TP_i}{\sum TP_i + \sum FN_i} \tag{5.8}$$

$$F1_{Mac} = \frac{2}{\frac{1}{Rec_{Mac}} + \frac{1}{Pre_{Mac}}} \tag{5.9}$$

$$F1_{Mic} = \frac{2}{\frac{1}{Rec_{Mic}} + \frac{1}{Pre_{Mic}}} \tag{5.10}$$

$$Spec_{Mac} = \frac{1}{4}\sum \frac{TN_i}{TN_i + FP_i} \tag{5.11}$$

$$Spec_{Mic} = \frac{\sum TN_i}{\sum TN_i + \sum FP_i} \tag{5.12}$$

where TP stands for True Positive, TN stands for True Negative, FP stands for False Positive, and FN stands for False Negative in the classification confusion matrix.

The quality of the response message was evaluated through expert scoring based on six criteria. Five criteria were chosen from widely adopted standards in the human evaluation of automatically generated text [167]: fluency, informativeness, relevance, grammaticality, and overall quality. Additionally, we introduced an "immersion" criterion to assess the level of distraction caused to pilots by correct identifications and the disruption caused by incorrect identifications, if such messages prompted out during flight. A lower score indicates a stronger negative impact from distractions or incorrect estimations, whereas a higher score indicates less negative impact. Response messages generated by GPT-4 from the VAM combination with the highest estimation accuracy were presented to five aviation experts who rated them using a 5-point Likert scale. The experts included two licensed pilots, two licensed ATCOs, and one aviation researcher. The ground truth, estimation result, and response message for the estimated positive cases were provided to the experts for evaluation. For cases where GPT-4 estimated the participant was conducting normal monitoring (Neg), no response messages were provided, and only the accuracy was given to the experts for reference in rating the scores.

5.3.3 Results

To establish a baseline, we first tested the performance using only the total fixation duration on each AOI. Corresponding to the 8x8 and 9x9 VAMs, this baseline performance was assessed using the total fixation durations for each AOI (*8-Sum*) and by adding a sum of all fixation durations (*9-Sum*). It provides less information on how the participant's attention transfers among these AOIs, which is equal to using solely the sum row in the 8x8 VAMs and 9x9 VAMs. Subsequently, the four VAM combinations were tested to identify the optimal format for estimation accuracy. The results are shown in Figure 5.6. Notably, it is a non-square confusion matrix with four rows and five columns. The 5th column represents the situation when the GPT-4 model estimates the participant was troubleshooting AOIs other than the three ground truths. As illustrated, using VAMs significantly improved estimation performance compared to solely using total fixation durations (*8-Sum* and *9-Sum*). Among the tested methods, *9-p* achieved the highest number of true positives (*157*).

Actual \ Prediction: **8-Sum**	Normal	ECAM &MFD	ND	LDG	Others
Normal	**67**	17	6	0	3
ECAM &MFD	5	**24**	0	0	2
ND	5	8	**8**	2	8
LDG	10	10	2	**4**	5

(a) Total fixation durations on AOIs (*8-Sum*)

Actual \ Prediction: **9-Sum**	Normal	ECAM &MFD	ND	LDG	Others
Normal	**64**	14	9	0	6
ECAM &MFD	5	**21**	3	0	2
ND	8	10	**9**	0	4
LDG	10	8	3	**5**	5

(b) Total fixation durations on AOIs + Sum duration (*9-Sum*)

Actual \ Prediction: **8-a**	Normal	ECAM &MFD	ND	LDG	Others
Normal	**77**	7	5	0	4
ECAM &MFD	0	**26**	2	0	3
ND	2	3	**23**	0	3
LDG	1	4	0	**19**	7

(c) 8*8 VAMs **without** P-value VAM (*8-a*)

Actual \ Prediction: **9-a**	Normal	ECAM &MFD	ND	LDG	Others
Normal	**77**	4	2	1	9
ECAM &MFD	1	**25**	2	0	3
ND	1	4	**23**	1	2
LDG	0	8	3	**15**	5

(d) 9*9 VAMs **without** P-value VAM (*9-a*)

Actual \ Prediction: **8-p**	Normal	ECAM &MFD	ND	LDG	Others
Normal	**77**	4	2	1	9
ECAM &MFD	1	**25**	2	0	3
ND	1	4	**23**	1	2
LDG	0	8	3	**15**	5

(c) 8*8 VAMs **with** P-value VAM (*8-p*)

Actual \ Prediction: **9-p**	Normal	ECAM &MFD	ND	LDG	Others
Normal	**80**	6	2	0	5
ECAM &MFD	1	**27**	1	0	2
ND	0	4	**23**	0	4
LDG	2	3	0	**22**	4

(d) 9*9 VAMs **with** P-value VAM (*9-p*)

Figure 5.6: Classification confusion matrix of different input

Based on the classification confusion matrix, results of macro-accuracy and micro-accuracy metrics were calculated and presented in Figure 5.7. The findings indicate that using VAM methods enhanced estimation performance across all four metrics, and that incorporating the P-value VAM to hint at the significant elements in the VAM further improved performance compared to treating all matrix elements equally.

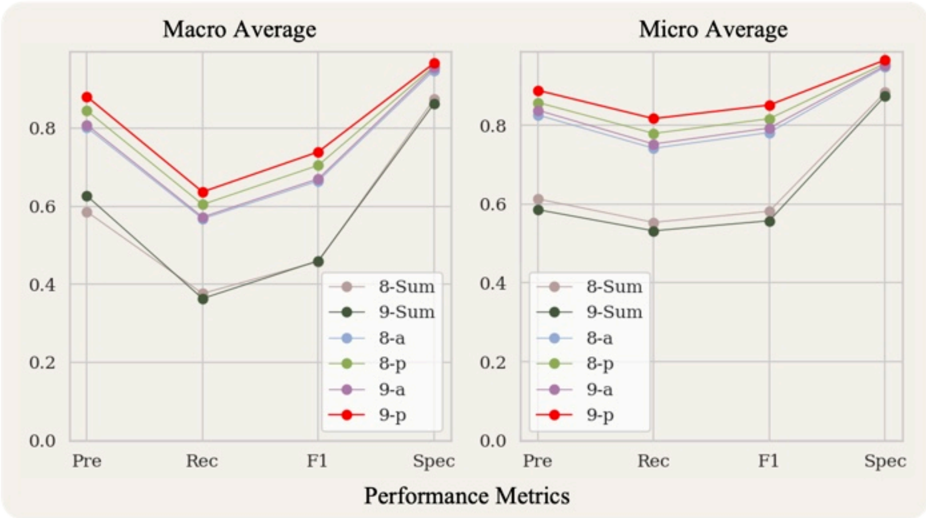

Figure 5.7: Macro and Micro evaluation metrics using different inputs

Specifically, *9-p* showed the best performance among the tested methods, achieving a macro average precision of *0.881*, recall of *0.637*, F1 score of *0.739*, and specificity of *0.966*. For the micro average, *9-p* achieved a precision of *0.889*, recall of *0.817*, F1 score of *0.852*, and specificity of *0.966*. The input using *8-a* and *9-a* demonstrated similar performance, with macro average precision around *0.80*, recall around *0.57*, F1 scores around *0.67*, and specificity around *0.95*. Their micro average precision were approximately *0.83*, recall were approximately *0.75*, F1 scores were approximately *0.78*, and specificity were approximately *0.95*. The performance metrics for *8-p* were higher than those for *8-a* and *9-a*, but still lower than for *9-p*.

The ground truth, estimation results, response messages, and accuracy metrics for *9-p* were provided to the experts for evaluation. 6.2 details the estimated positive cases, while negative cases (normal monitoring) were excluded as they did not generate response messages. The expert evaluation results are summarized in Table 5.1. All experts rated *5* for grammaticality and agreed that the GPT-generated messages had good overall quality, with an average score of *4.4*. The average scores for fluency and relevance were also *4.4*. However, the scores for informativeness and immersion were lower, with an average of *3.6*, indicating some concerns in these aspects.

Table 5.1: Expert scoring result

Expert	Fluency	Informa-tiveness	Relevance	Gramma-ticality	Immersion	Overall Quality
Pilot 1	5	4	5	5	3	5
Pilot 2	4	4	4	5	5	4
ATCO 1	4	3	4	5	3	4
ATCO 2	5	4	5	5	4	5
Researcher	4	3	4	5	3	4
Average	4.4	3.6	4.4	5	3.6	4.4

5.4 Discussion

When pilots are actively engaged in the loop (ITL), timely automation support can facilitate them to acquire information more effectively and troubleshoot incidents more efficiently. This study proposes a method to identify the troubleshooting of pilots by tokenizing the eye-tracking data into Visual Attention Matrices (VAMs) and integrating with LLMs. A case study validates the effectiveness of this approach in identifying pilots' troubleshooting and proposing user-friendly support. Comparing different VAM configurations, the use of 9x9 VAMs with empirical P-Value VAMs (*9-p*) achieved optimal estimation accuracy, with F1 scores of 0.739 for Macro-average and 0.852 for Micro-average. Expert scoring confirmed the acceptance of using GPT-4 to proactively propose support to pilots. The implications derived from these results and the study's limitations are discussed below.

5.4.1 Implications derived from the results

Firstly, comparing the estimation performance using VAMs against baseline methods (solely total fixation on each AOI) revealed significant enrichment in understanding pilots' visual attention distribution through AOI transitions. Though the individual saccade behaviors represent the original visual transitions, the VAM method summarises these transitions to reflect the trend rather than maintaining individual saccades to avoid the load of massive data processing. This is in line with the previous methods such as gaze transition entropy [168, 169] and AOI Rivers [170]. The summarization mitigates the challenge of LLMs' deficiency in processing massive physiological data with no semantic information.

Secondly, VAMs including sum rows and columns (9x9 VAMs) outperformed those without (8x8 VAMs), suggesting that simple features might also enhance LLM's reasoning ability. Moreover, the use of empirical P-value VAMs outperformed its absence, underscoring the importance of highlighting the critical features for the LLM. These findings highlight the significance of feature engineering based on empirical knowledge [171], especially for the industrial applications of large models. For example, Figure 5.6 demonstrated that the ECAM& MFD had higher TP and NP than LDG, potentially caused by the nature that ECAM& MFD provide more critical information than LDG and intrinsically received more attention. A guess can be made that if this empirical knowledge was provided to the GPT-4 model, the estimation accuracy on the LDG might be improved.

Thirdly, all the experts rated 4 or above in fluency, relevance, grammaticality, and overall quality for the GPT-generated response messages. This indicates overall acceptance of the method in identifying the troubleshooting activity and text quality. However, concerns were raised about informativeness and immersion criteria, suggesting a need for more granular yet concise output from GPT-generated messages. Future efforts should focus on the deeper integration of LLMs with in-flight tasks and advancing human-AI teaming paradigms.

5.4.2 Limitations

This study involved nineteen pilots with varying experience levels and flight hours, which may influence cockpit visual scanning behaviors. Fortunately, distinct visual patterns captured by our method led to satisfactory estimation performance. In the future, it would be ideal to separate the pilots of different experience levels into different groups of the same size to further study experience-related eye movement characteristics. Furthermore, the ground truth of the classification in this study was simplified and could be biased. For example, though the error message was shown on designed instruments, the pilot may feel a need to acquire information from other instruments for cross-validation. This might impact the label of the data and the result of accuracy calculation. Therefore, it is of great value to improve the experiment design to label the pilots' spontaneous troubleshooting without interrupting their immersive operations in-flight tasks.

In summary, the results demonstrated that the proposed method effectively identified the pilots' troubleshooting and had the ability to propose user-friendly support. It fills the critical gap in detecting and actively responding to the needs of the ITL operators. This methodology provides a novel format for eye-tracking data processing, and serves as a base for future HCI studies in the new era of large models.

5.5 Concluding remarks

Optimizing Human-Computer Interaction (HCI) in highly automated cockpits is crucial for aviation safety. While numerous studies have focused on mitigating risks associated with the OOTL phenomenon, a critical gap remains regarding the need for support of In-the-Loop (ITL) pilots. To address this gap, our study introduced a method to tokenize eye-tracking data and integrate it with Large Language Models (LLMs) to identify and respond to ITL pilots' troubleshooting.

This research makes three significant contributions to HCI in aviation. First, it presents a novel method for tokenizing pilots' eye-tracking data into Visual Attention Matrices (VAMs), which capture visual attention transitions across cockpit instruments and their distribution. The VAM effectively summarizes the characteristics of pilots' normal monitoring and troubleshooting behaviours, providing a suitable format for LLM-based processing and reasoning. Second, the study employs three empirical VAMs (Normal Range VAM, Abnormal Range VAM, and P-Value VAM) as empirical knowledge to support LLM reasoning. Compared to training and tuning LLMs with all individual VAMs, the use of empirical

VAMs significantly reduces resource costs. Third, beyond the identification of the troubleshooting behaviour, the study proposes actively providing support using the natural interaction capability of LLM. This approach offers insights for future LLM integration in aviation systems and establishes a foundation for HCI in the new era of automation.

Despite limitations related to participants' experience variability and the unavoidable biases in ground truth labelling, the results demonstrate that the proposed method can effectively capture the characteristics of pilots' eye movements to discern different monitoring behaviours. Future research aims to achieve a more balanced participant composition by categorizing pilots based on experience levels and ensuring equal group sizes. Furthermore, integrating more aviation-specific knowledge into the method will improve estimation accuracy and optimize the generated interaction messages. This will enable a more comprehensive understanding of pilots' needs during ITL status and advance HCI to enhance aviation safety.

Chapter 6

Conclusion

This chapter concludes this work, with the key findings and main contributions summarised in Section 6.1. Meanwhile, the limitations and future research directions of this study are discussed in Section 6.2.

6.1 Contribution

This thesis advances the understanding and application of human factors and automation in aviation through a systematic exploration of visual attention-based methodologies. The work spans the development of predictive models, frameworks for explainable status identification, and innovative interaction designs. Key contributions are outlined as follows:

1. **Development of the Flashlight model**

   The Flashlight model combines attention distribution and resource metrics, refining traditional gaze metrics by emphasizing specific Areas of Interest (AOIs) critical to performance prediction.

SHAP analysis is used to identify pivotal eye-tracking measurements, enriching interpretability and enabling more targeted applications in aviation safety.

Integrated the classic cognitive theories, such as James' Spotlight and Wickens' information processing models, into practical applications for aviation. It advances the development of novel theories to optimise aviation safety.

2. **Introduction of the explainable OOTL identification framework**

   The Visual Attention Traces (VATs) innovatively encode the eye-tracking data, capturing both temporal and spatial dynamics. This novel method enables robust characterization of pilots' behaviors.

   The VALIO framework applies LTL methods to generate human-readable formulas, improving explainability in identifying OOTL status.

   The framework demonstrates superior performance in shorter time windows, highlighting its practical applicability for real-time aviation operations.

   The obtained LTL_f formulas in this study directly provide insights into pilots' OOTL status and contribute to the understanding of pilots' visual behaviours.

3. **Integration of LLMs for ITL Pilot Support**

   This work develops a Visual Attention Matrices (VAMs) method to tokenize eye-tracking data into formats suitable for LLM processing. It effectively summarizes the characteristics of normal and troubleshooting behaviours.

   The method utilises empirical VAMs to incorporate empirical knowledge for LLM reasoning and interaction. It innovatively reduces the resource costs of training and tuning LLMs.

By integrating eye-tracking and LLM methods, this work proposes a proactive HCI with natural interaction capabilities. It leverages the capability of LLMs to provide timely and context-aware support, advancing HCI in highly automated cockpits.

In summary, this thesis contributes to the fields of *human-in-the-loop HCI in cockpit to enhance aviation safety by providing innovative methods for monitoring, predicting, and supporting pilot performance in both OOTL and ITL scenarios*. Meanwhile, the methods developed and the knowledge obtained from this work might also be adopted in other visual-based tasks where human operators act as monitor roles, such as drivers of autonomous vehicles and Vessel Traffic Controllers.

6.2 Limitations and future research

In this thesis, eye-tracking data are innovatively encoded to integrate the attention resources perspective and attention distribution perspective. The applications of LTL_f methods and LLMs are leveraged, which are relatively novel attempts without broad explorations. Although a great effort was put forth with regard to the and feature engineering and data analysis, this thesis has inevitable limitations and constraints as discussed in the following:

1. *Participant Experience*: Some of the studies involved student pilots and simplified scenarios, limiting the generalizability of findings to more complex and realistic aviation environments.

2. *Binary Classification of Pilot Status*: It weakened some nuanced reality of pilot behaviours and interactions with automation.

3. *Ground Truth Labelling Bias*: The reliance on labelled data introduces potential biases that may affect model performance and interpretation accuracy.

4. *Static settings of AOIs*: This study defined AOIs statically based on the static infrastructure in the cockpit, limiting the applications to the scenarios where the AOIs are dynamic (e.g., Air Traffic Control).

5. *Simplified Experimental Tasks*: Laboratory conditions and predefined tasks restrict the applicability of findings to dynamic and unpredictable real-world aviation scenarios.

Based on these limitations and the remaining research gaps, some future research directions are discussed.

1. *Expanding Participant Profiles*: Future studies should involve more licensed pilots and encompass diverse expertise levels to ensure more representative findings.

2. *Enhancing Experimental Realism*: Future studies should try to establish real-world flight scenarios with dynamic variables to better validate the proposed methods.

3. *Refinement of Classification Models*: Future studies will incorporate more detailed degrees of OOTL and ITL, along with hybrid states. Meanwhile,

other Computer Vision techniques will be introduced to support a real-time AOI definition with dynamic visual targets.

4. *Broader Range of Eye-Tracking Metrics*: Future studies will explore additional metrics. The other eye-tracking metrics, such as gaze entropy and blinks, will be utilised to offer deeper insights into cognitive processes.

5. *Advanced Interaction Models*: Future studies will leverage LLMs for multi-modal integration (e.g., combining audio and physiological signals). The LLM will be customised with multi-modal information sources to enhance the accuracy and responsiveness of automation support.

References

[1] Wiegmann, D. A., Shappell, S. A. “Human Error Analysis of Commercial Aviation Accidents Using the Human Factors Analysis and Classification System (HFACS)”. In: (DOT/FAA/AM-01/3, Feb. 1, 2001). URL: https://rosap.ntl.bts.gov/view/dot/15409.

[2] Kharoufah, H., Murray, J., Baxter, G., Wild, G. “A review of human factors causations in commercial air transport accidents and incidents: From to 2000–2016”. In: *Progress in Aerospace Sciences* 99 (May 2018), pp. 1–13. ISSN: 03760421. DOI: 10.1016/j.paerosci.2018.03.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S0376042117302154 (visited on 02/29/2024).

[3] Sant’Anna, D. A. L. M. D., Hilal, A. V. G. D. “The impact of human factors on pilots’ safety behavior in offshore aviation companies: A brazilian case”. In: *Safety Science* 140 (Aug. 2021), p. 105272. ISSN: 09257535. DOI: 10.1016/j.ssci.2021.105272. URL: https://linkinghub.elsevier.com/retrieve/pii/S092575352100117X (visited on 02/29/2024).

[4] Osunwusi, A. O. “Aviation Automation and CNS/ATM-related Human-Technology Interface: ATSEP Competency Considerations”. In: *International Journal of Aviation, Aeronautics, and Aerospace* (2019). ISSN: 23746793. DOI: 10.15394/ijaaa.2019.1390. URL: https://commons.erau.edu/ijaaa/vol6/iss4/13/ (visited on 10/04/2024).

[5] Council, N. R. *Taking Flight: Education and Training for Aviation Careers*. Ed. by Janet S. Hansen and Jr. Clinton V. Oster. Washington, DC: The National Academies Press, 1997. ISBN: 978-0-309-05676-2. DOI: 10.17226/5433. URL: https://nap.nationalacademies.org/catalog/5433/taking-flight-education-and-training-for-aviation-careers.

[6] Kaber, D. B. “Issues in Human–Automation Interaction Modeling: Presumptive Aspects of Frameworks of Types and Levels of Automation”. In: *Journal of Cognitive Engineering and Decision Making* 12.1 (Mar. 2018), pp. 7–24. ISSN: 1555-3434. DOI: 10.1177/1555343417737203. URL: https://journals.sagepub.com/doi/10.1177/1555343417737203 (visited on 10/04/2024).

[7] Endsley, M. R. “Level of Automation Forms a Key Aspect of Autonomy Design”. In: *Journal of Cognitive Engineering and Decision Making* 12.1 (Mar. 2018), pp. 29–34. ISSN: 1555-3434. DOI: 10.1177/1555343417723432. URL: https://journals.sagepub.com/doi/10.1177/1555343417723432 (visited on 10/04/2024).

[8] Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., Green, C. A., Katazaki, S., Monk, C., Itoh, M., McGehee, D.,

Sunda, T., Unoura, K., Victor, T., Schieben, A., Keinath, A. "The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications". In: *Cognition, Technology & Work* 21.1 (Feb. 2019), pp. 87–98. ISSN: 1435-5558, 1435-5566. DOI: 10.1007/s10111-018-0525-8. URL: http://link.springer.com/10.1007/s10111-018-0525-8 (visited on 01/12/2023).

[9] Ivanov, S. H. "Automated decision-making". In: *foresight* 25.1 (Mar. 16, 2023), pp. 4–19. ISSN: 1463-6689, 1463-6689. DOI: 10.1108/FS-09-2021-0183. URL: https://www.emerald.com/insight/content/doi/10.1108/FS-09-2021-0183/full/html (visited on 10/04/2024).

[10] Endsley, M. R., Kiris, E. O. "The Out-of-the-Loop Performance Problem and Level of Control in Automation". In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37.2 (June 1995), pp. 381–394. ISSN: 0018-7208, 1547-8181. DOI: 10.1518/001872095779064555. (Visited on 03/03/2023).

[11] **Lyu, M.**, Li, F., Lee, C.-H., Chen, C.-H. "VALIO: Visual attention-based linear temporal logic method for explainable out-of-the-loop identification". In: *Knowledge-Based Systems* 299 (Sept. 2024), p. 112086. ISSN: 09507051. DOI: 10.1016/j.knosys.2024.112086. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950705124007202 (visited on 06/22/2024).

[12] Gerbert, K., Kemmler, R. "The causes of causes: determinants and background variables of human factor incidents and accidents". In: *Ergonomics* 29.11 (Nov. 1986), pp. 1439–1453. ISSN: 0014-0139, 1366-5847. DOI:

10.1080/00140138608967257. URL: http://www.tandfonline.com/doi/abs/10.1080/00140138608967257 (visited on 11/08/2023).

[13] Mosier, K. L., Skitka, L. J., Korte, K. J., Mouloua, M., Parasuraman, R. “Cognitive and social psychological issues in flight crew/automation interaction”. In: *Human performance in automated systems: Current research and trends* (1994). Publisher: Erlbaum Hillsdale, NJ, pp. 191–197.

[14] Alharasees, O., Adali, O. H., Kale, U. “Comprehensive Review on Aviation Operator’s Total Loads”. In: *2022 New Trends in Aviation Development (NTAD)*. 2022 New Trends in Aviation Development (NTAD). Novy Smokovec, Slovakia: IEEE, Nov. 24, 2022, pp. 15–20. ISBN: 9798350320367. DOI: 10.1109/NTAD57912.2022.10013609. URL: https://ieeexplore.ieee.org/document/10013609/ (visited on 11/07/2024).

[15] Ghaderi, M., Ben Abdessalem, H., Frasson, C. “An Analysis of Mental Workload Involved in Piloting Tasks”. In: *Novel & Intelligent Digital Systems: Proceedings of the 2nd International Conference (NiDS 2022)*. Ed. by Akrivi Krouska, Christos Troussas, and Jaime Caro. Vol. 556. Series Title: Lecture Notes in Networks and Systems. Cham: Springer International Publishing, 2023, pp. 211–220. ISBN: 978-3-031-17600-5 978-3-031-17601-2. DOI: 10.1007/978-3-031-17601-2_21. URL: https://link.springer.com/10.1007/978-3-031-17601-2_21 (visited on 11/07/2024).

[16] Hart, S. G., Staveland, L. E. “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research”. In: *Advances in Psychology*. Vol. 52. Elsevier, 1988, pp. 139–183. ISBN: 978-0-444-70388-

0. DOI: 10.1016/S0166-4115(08)62386-9. URL: https://linkinghub.elsevier.com/retrieve/pii/S0166411508623869 (visited on 03/05/2024).

[17] Frank, N., Le, P., Mills, E., Davis, K. G. "Micromovements and discomfort associated with flight mission with helmet operation tasks with different levels of cognitive workload". In: *International Journal of Industrial Ergonomics* 95 (May 2023), p. 103441. ISSN: 01698141. DOI: 10.1016/j.ergon.2023.103441. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169814123000331 (visited on 06/12/2024).

[18] Endsley, M. "Situation awareness global assessment technique (SAGAT)". In: *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*. IEEE 1988 National Aerospace and Electronics Conference. Dayton, OH, USA: IEEE, 1988, pp. 789–795. DOI: 10.1109/NAECON.1988.195097. URL: http://ieeexplore.ieee.org/document/195097/ (visited on 03/05/2024).

[19] Van Benthem, K., Brightman, K., Riguero, E., Herdman, C. M. "Results and methodology for classifying high risk pilots using CANFLY: A cognitive health screening tool for aviators". In: *International Journal of Industrial Ergonomics* 101 (May 2024), p. 103578. ISSN: 01698141. DOI: 10.1016/j.ergon.2024.103578. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169814124000349 (visited on 06/12/2024).

[20] Yiu, C. Y., Ng, K. K., Li, X., Zhang, X., Li, Q., Lam, H. S., Chong, M. H. "Towards safe and collaborative aerodrome operations: Assessing shared situational awareness for adverse weather detection with EEG-enabled Bayesian neural networks". In: *Advanced Engineering In-*

formatics 53 (Aug. 2022), p. 101698. ISSN: 14740346. DOI: 10.1016/j.aei.2022.101698. URL: https://linkinghub.elsevier.com/retrieve/pii/S1474034622001574 (visited on 01/30/2023).

[21] Li, Q., Ng, K. K., Yiu, C. Y., Yuan, X., So, C. K., Ho, C. C. "Securing air transportation safety through identifying pilot's risky VFR flying behaviours: An EEG-based neurophysiological modelling using machine learning algorithms". In: *Reliability Engineering & System Safety* 238 (Oct. 2023), p. 109449. ISSN: 09518320. DOI: 10.1016/j.ress.2023.109449. URL: https://linkinghub.elsevier.com/retrieve/pii/S0951832023003630 (visited on 05/22/2024).

[22] Hamann, A., Carstengerdes, N. "Assessing the development of mental fatigue during simulated flights with concurrent EEG-fNIRS measurement". In: *Scientific Reports* 13.1 (Mar. 23, 2023), p. 4738. ISSN: 2045-2322. DOI: 10.1038/s41598-023-31264-w. URL: https://www.nature.com/articles/s41598-023-31264-w (visited on 11/08/2023).

[23] Pan, T., Wang, H., Si, H., Liu, H., Xu, M. "Research on the Identification of Pilots' Fatigue Status Based on Functional Near-Infrared Spectroscopy". In: *Aerospace* 9.3 (Mar. 21, 2022), p. 173. ISSN: 2226-4310. DOI: 10.3390/aerospace9030173. URL: https://www.mdpi.com/2226-4310/9/3/173 (visited on 02/29/2024).

[24] Qin, H., Zhou, X., Ou, X., Liu, Y., Xue, C. "Detection of mental fatigue state using heart rate variability and eye metrics during simulated flight". In: *Human Factors and Ergonomics in Manufacturing & Service*

Industries 31.6 (Nov. 2021), pp. 637–651. ISSN: 1090-8471. DOI: 10.1002/hfm.20927.

[25] White, A., O'Hare, D. "In plane sight: Inattentional blindness affects visual detection of external targets in simulated flight". In: *Applied Ergonomics* 98 (Jan. 2022), p. 103578. ISSN: 00036870. DOI: 10.1016/j.apergo.2021.103578. URL: https://linkinghub.elsevier.com/retrieve/pii/S0003687021002258 (visited on 01/26/2022).

[26] Li, Z., Li, R., Yuan, L., Cui, J., Li, F. "A benchmarking framework for eye-tracking-based vigilance prediction of vessel traffic controllers". In: *Engineering Applications of Artificial Intelligence* 129 (Mar. 2024), p. 107660. ISSN: 09521976. DOI: 10.1016/j.engappai.2023.107660. URL: https://linkinghub.elsevier.com/retrieve/pii/S0952197623018444 (visited on 12/15/2023).

[27] Pütz, S., Mertens, A., Chuang, L., Nitsch, V. "Physiological measures of operators' mental state in supervisory process control tasks: a scoping review". In: *Ergonomics* (Dec. 21, 2023), pp. 1–30. ISSN: 0014-0139, 1366-5847. DOI: 10.1080/00140139.2023.2289858. URL: https://www.tandfonline.com/doi/full/10.1080/00140139.2023.2289858 (visited on 03/05/2024).

[28] Wickens, C. "Attention: Theory, Principles, Models and Applications". In: *International Journal of Human–Computer Interaction* 37.5 (Mar. 16, 2021), pp. 403–417. ISSN: 1044-7318, 1532-7590. DOI: 10.1080/10447318.2021.1874741. URL: https://www.tandfonline.com/doi/full/10.1080/10447318.2021.1874741 (visited on 11/04/2023).

[29] Mahanama, B., Jayawardana, Y., Rengarajan, S., Jayawardena, G., Chukoskie, L., Snider, J., Jayarathna, S. "Eye Movement and Pupil Measures: A Review". In: *Frontiers in Computer Science* 3 (Jan. 11, 2022), p. 733531. ISSN: 2624-9898. DOI: 10.3389/fcomp.2021.733531. URL: https://www.frontiersin.org/articles/10.3389/fcomp.2021.733531/full (visited on 11/09/2022).

[30] Kang, Y., Liu, F., Chen, W., Li, X., Tao, Y., Huang, W. "Recognizing situation awareness of forklift operators based on eye-movement & EEG features". In: *International Journal of Industrial Ergonomics* 100 (Mar. 2024), p. 103552. ISSN: 01698141. DOI: 10.1016/j.ergon.2024.103552. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169814124000088 (visited on 03/04/2024).

[31] Peißl, S., Wickens, C. D., Baruah, R. "Eye-Tracking Measures in Aviation: A Selective Literature Review". In: *The International Journal of Aerospace Psychology* 28.3 (Oct. 2, 2018), pp. 98–112. ISSN: 2472-1840, 2472-1832. DOI: 10.1080/24721840.2018.1514978. URL: https://www.tandfonline.com/doi/full/10.1080/24721840.2018.1514978 (visited on 02/07/2022).

[32] Li, F., Xu, G., Feng, S. "Eye Tracking Analytics for Mental States Assessment – A Review*". In: *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Melbourne, Australia: IEEE, Oct. 17, 2021, pp. 2266–2271. ISBN: 978-1-66544-207-7. DOI:

10.1109/SMC52423.2021.9658674. URL: https://ieeexplore.ieee.org/document/9658674/ (visited on 01/10/2022).

[33] **Lyu, M.**, Fan, L., Gangyan, X., Su, H. “Leveraging eye-tracking technologies to promote aviation safety- A review of key aspects, challenges, and future perspectives”. In: *Safety Science* 168 (2023), p. 106295. ISSN: 0925-7535. DOI: https://doi.org/10.1016/j.ssci.2023.106295. URL: https://www.sciencedirect.com/science/article/pii/S0925753523002370.

[34] Li, Q., Ng, K. K. H., Yu, S. C. M., Yiu, C. Y., **Lyu, M.** “Recognising situation awareness associated with different workloads using EEG and eye-tracking features in air traffic control tasks”. In: *Knowledge-Based Systems* 260 (2023), p. 110179. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2022.110179. URL: https://www.sciencedirect.com/science/article/pii/S0950705122012758.

[35] Cui, P., Athey, S. “Stable learning establishes some common ground between causal inference and machine learning”. In: *Nature Machine Intelligence* 4.2 (Feb. 23, 2022), pp. 110–115. ISSN: 2522-5839. DOI: 10.1038/s42256-022-00445-z. URL: https://www.nature.com/articles/s42256-022-00445-z (visited on 09/07/2023).

[36] Saeed, W., Omlin, C. “Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities”. In: *Knowledge-Based Systems* 263 (Mar. 2023), p. 110273. ISSN: 09507051. DOI: 10.1016/j.knosys.2023.110273. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950705123000230 (visited on 02/21/2024).

[37] Li, F., Chen, C.-H., Lee, C.-H., Khoo, L.-P. “A User Requirement-driven Approach Incorporating TRIZ and QFD for Designing a Smart Vessel Alarm System to Reduce Alarm Fatigue”. In: *Journal of Navigation* 73.1 (Jan. 2020), pp. 212–232. ISSN: 0373-4633, 1469-7785. DOI: 10.1017/S0373463319000547. URL: https://www.cambridge.org/core/product/identifier/S0373463319000547/type/journal_article (visited on 12/14/2021).

[38] Li, F., Lee, C.-H., Xu, G., Chen, C.-H., Khoo, L. P. “A QFD-enabled conceptualization for reducing alarm fatigue in vessel traffic service centre”. In: *Transdisciplinary Engineering: A Paradigm Shift*. IOS Press, 2017, pp. 821–828.

[39] Camacho, A., McIlraith, S. A. “Learning Interpretable Models Expressed in Linear Temporal Logic”. In: *Proceedings of the International Conference on Automated Planning and Scheduling* 29 (May 25, 2021), pp. 621–630. ISSN: 2334-0843, 2334-0835. DOI: 10.1609/icaps.v29i1.3529. URL: https://ojs.aaai.org/index.php/ICAPS/article/view/3529 (visited on 11/09/2023).

[40] Faulhaber, A. K., Friedrich, M., Kapol, T. “Absence of Pilot Monitoring Affects Scanning Behavior of Pilot Flying: Implications for the Design of Single-Pilot Cockpits”. In: *Human Factors* 64.2 (2022), pp. 278–290. DOI: 10.1177/0018720820939691.

[41] Avril, E. “Providing different levels of accuracy about the reliability of automation to a human operator: impact on human performance”. In: *Ergonomics* 66.2 (Feb. 1, 2023), pp. 217–226. ISSN: 0014-0139, 1366-5847.

DOI: 10.1080/00140139.2022.2069870. URL: https://www.tandfonline.com/doi/full/10.1080/00140139.2022.2069870 (visited on 10/04/2024).

[42] Ban, G., Park, W. "Effects of In-Vehicle Touchscreen Location on Driver Task Performance, Eye Gaze Behavior, and Workload During Conditionally Automated Driving: Nondriving-Related Task and Take-Over". In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* (Jan. 19, 2024), p. 00187208241226838. ISSN: 0018-7208, 1547-8181. DOI: 10.1177/00187208241226838. (Visited on 03/28/2024).

[43] Lorenz, B., Di Nocera, F., Röttger, S., Parasuraman, R. "The Effects of Level of Automation on the Out-of-the-Loop Unfamiliarity in a Complex Dynamic Fault-Management Task during Simulated Spaceflight Operations". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 45.2 (Oct. 2001), pp. 44–48. ISSN: 2169-5067, 1071-1813. DOI: 10.1177/154193120104500209. URL: http://journals.sagepub.com/doi/10.1177/154193120104500209 (visited on 11/23/2023).

[44] Gouraud, J., Delorme, A., Berberian, B. "Autopilot, Mind Wandering, and the Out of the Loop Performance Problem". In: *Frontiers in Neuroscience* 11 (Oct. 5, 2017), p. 541. ISSN: 1662-453X. DOI: 10.3389/fnins.2017.00541. URL: http://journal.frontiersin.org/article/10.3389/fnins.2017.00541/full (visited on 07/20/2023).

[45] Louw, T., Merat, N. "Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation". In: *Transportation Research Part C: Emerging Technologies* 76 (Mar. 2017), pp. 35–

50. ISSN: 0968090X. DOI: 10.1016/j.trc.2017.01.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0968090X17300013 (visited on 02/01/2023).

[46] Sebok, A., Wickens, C. D. "Implementing Lumberjacks and Black Swans Into Model-Based Tools to Support Human–Automation Interaction". In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 59.2 (Mar. 2017), pp. 189–203. ISSN: 0018-7208, 1547-8181. DOI: 10.1177/0018720816665201. (Visited on 02/01/2023).

[47] Li, Q., Ng, K. K., Chu, S. T., Lau, T. Y., Leung, C. H. "Revealing the effects of increased workload and distraction on the pilot's situation awareness neurobehavioral activities". In: *AIAA AVIATION 2023 Forum*. AIAA AVIATION 2023 Forum. San Diego, CA and Online: American Institute of Aeronautics and Astronautics, June 12, 2023. ISBN: 978-1-62410-704-7. DOI: 10.2514/6.2023-3407. URL: https://arc.aiaa.org/doi/10.2514/6.2023-3407 (visited on 06/12/2024).

[48] Conati, C., Merten, C. "Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation". In: *Knowledge-Based Systems* 20.6 (Aug. 2007), pp. 557–574. ISSN: 09507051. DOI: 10.1016/j.knosys.2007.04.010. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950705107000469 (visited on 01/05/2022).

[49] Schnebelen, D., Charron, C., Mars, F. "Estimating the out-of-the-loop phenomenon from visual strategies during highly automated driving". In: *Accident Analysis & Prevention* 148 (Dec. 2020), p. 105776. ISSN: 00014575. DOI: 10.1016/j.aap.2020.105776. URL: https:

//linkinghub.elsevier.com/retrieve/pii/S0001457520315967 (visited on 11/26/2023).

[50] Di Stasi, L. L., Diaz-Piedra, C. "Re-examining the Pioneering Studies on Eye Movements in Aviation: Connecting the Past to the Present". In: *The International Journal of Aerospace Psychology* 31.2 (Apr. 3, 2021), pp. 122–134. ISSN: 2472-1832. DOI: 10.1080/24721840.2020.1862658.

[51] Griffith, C. "The organic effects of repeated bodily rotation." In: *Journal of Experimental Psychology* 3 (Feb. 1920), pp. 15–46. DOI: 10.1037/h0069767.

[52] Vasigh, B., Azadian, F. "The Globalization and Evolution of the Aviation Industry". In: Mar. 2022, pp. 1–65. ISBN: 978-3-030-82449-5. DOI: 10.1007/978-3-030-82450-1_1.

[53] Bhatia, P., Khosla, A., Singh, G. "A Review on Eye Tracking Technology:" in: *Advances in Medical Diagnosis, Treatment, and Care*. Ed. by Tanu Wadhera and Deepti Kakkar. IGI Global, 2020, pp. 107–130. ISBN: 978-1-79983-069-6 978-1-79983-070-2. DOI: 10.4018/978-1-7998-3069-6.ch007. URL: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-7998-3069-6.ch007 (visited on 04/20/2023).

[54] Klaib, A. F., Alsrehin, N. O., Melhem, W. Y., Bashtawi, H. O., Magableh, A. A. "Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies". In: *Expert Systems with Applications* 166 (2021), p. 114037. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2020.114037.

URL: https://www.sciencedirect.com/science/article/pii/S0957417420308071.

[55] Ziv, G. "Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation". In: *The International Journal of Aviation Psychology* 26.3 (2017), pp. 75–104. ISSN: 1050-8414. DOI: 10.1080/10508414.2017.1313096.

[56] Hansen, D. W., Ji, Q. "In the Eye of the Beholder: A Survey of Models for Eyes and Gaze". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.3 (2010), pp. 478–500. DOI: 10.1109/TPAMI.2009.30. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-77949331734&doi=10.1109%2fTPAMI.2009.30&partnerID=40&md5=19f5763a890a51fc4e87d640a724abd8.

[57] Majaranta, P., Bulling, A. "Eye Tracking and Eye-Based Human–Computer Interaction". In: *Advances in Physiological Computing*. Ed. by Stephen H. Fairclough and Kiel Gilleade. London: Springer London, 2014, pp. 39–65. ISBN: 978-1-4471-6392-3. DOI: 10.1007/978-1-4471-6392-3_3. URL: https://doi.org/10.1007/978-1-4471-6392-3_3.

[58] Sarkar, A., Sanyal, G., Majumder, S. "Performance evaluation of an eye tracking system under varying conditions". In: *International Journal of Computer Science and Network Security (IJCSNS)* 17.4 (2017), p. 182. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85092394235&partnerID=40&md5=53246842c7b5d9aaf18244444c54b83b.

[59] Diaz-Piedra, C., Rieiro, H., Cherino, A., Fuentes, L. J., Catena, A., Di Stasi, L. L. "The effects of flight complexity on gaze entropy: An

experimental study with fighter pilots". In: *Applied Ergonomics* 77 (May 2019), pp. 92–99. ISSN: 00036870. DOI: 10.1016/j.apergo.2019.01.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S0003687019300274 (visited on 01/26/2022).

[60] Glöckner, A., Herbold, A.-K. "An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes". In: *Journal of Behavioral Decision Making* 24.1 (Jan. 2011), pp. 71–98. ISSN: 08943257. DOI: 10.1002/bdm.684. URL: https://onlinelibrary.wiley.com/doi/10.1002/bdm.684 (visited on 03/11/2022).

[61] Li, F., Chen, C.-H., Lee, C.-H., Feng, S. "Artificial intelligence-enabled non-intrusive vigilance assessment approach to reducing traffic controller's human errors". In: *Knowledge-Based Systems* (Dec. 2021), p. 108047. ISSN: 09507051. DOI: 10.1016/j.knosys.2021.108047. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950705121011369 (visited on 01/03/2022).

[62] Li, F., Lee, C.-H., Feng, S., Trappey, A., Gilani, F. "Prospective on Eye-Tracking-based Studies in Immersive Virtual Reality". In: *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD). Dalian, China: IEEE, May 5, 2021, pp. 861–866. ISBN: 978-1-72816-597-4. DOI: 10.1109/CSCWD49262.2021.9437692. URL: https://ieeexplore.ieee.org/document/9437692/ (visited on 12/14/2021).

[63] Li, Q., Ng, K. K., Fan, Z., Yuan, X., Liu, H., Bu, L. “A human-centred approach based on functional near-infrared spectroscopy for adaptive decision-making in the air traffic control environment: A case study”. In: *Advanced Engineering Informatics* 49 (Aug. 2021), p. 101325. ISSN: 14740346. DOI: 10.1016/j.aei.2021.101325. URL: https://linkinghub.elsevier.com/retrieve/pii/S1474034621000781 (visited on 02/02/2023).

[64] Li, X., Zheng, P., Bao, J., Gao, L., Xu, X. “Achieving Cognitive Mass Personalization via the Self-X Cognitive Manufacturing Network: An Industrial Knowledge Graph- and Graph Embedding-Enabled Pathway”. In: *Engineering* (Oct. 2021), S2095809921004252. ISSN: 20958099. DOI: 10.1016/j.eng.2021.08.018. URL: https://linkinghub.elsevier.com/retrieve/pii/S2095809921004252 (visited on 12/16/2022).

[65] Bruder, C., Hasse, C. “Differences between experts and novices in the monitoring of automated systems”. In: *International Journal of Industrial Ergonomics* 72 (July 2019), pp. 1–11. ISSN: 0169-8141. DOI: 10.1016/j.ergon.2019.03.007.

[66] Jin, H., Hu, Z., Li, K., Chu, M., Zou, G., Yu, G., Zhang, J. “Study on How Expert and Novice Pilots Can Distribute Their Visual Attention to Improve Flight Performance”. In: *IEEE Access* 9 (2021), pp. 44757–44769. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3066880. URL: https://ieeexplore.ieee.org/document/9380452/ (visited on 06/17/2022).

[67] Di Stasi, L. L., McCamy, M. B., Martinez-Conde, S., Gayles, E., Hoare, C., Foster, M., Catena, A., Macknik, S. L. “Effects of long and short simulated flights on the saccadic eye movement velocity of aviators”. In: *Physiology & Behavior* 153 (Jan. 1, 2016), pp. 91–96. ISSN: 0031-9384. DOI: 10.1016/j.physbeh.2015.10.024.

[68] Diaz-Piedra, C., Rieiro, H., Suarez, J., Rios-Tejada, F., Catena, A., Luigi Di Stasi, L. “Fatigue in the military: towards a fatigue detection test based on the saccadic velocity”. In: *Physiological Measurement* 37.9 (Sept. 2016), N62–N75. ISSN: 0967-3334. DOI: 10.1088/0967-3334/37/9/N62.

[69] Naeeri, S., Kang, Z., Mandal, S., Kim, K. “Multimodal Analysis of Eye Movements and Fatigue in a Simulated Glass Cockpit Environment”. In: *Aerospace* 8.10 (Oct. 1, 2021), p. 283. ISSN: 2226-4310. DOI: 10.3390/aerospace8100283. URL: https://www.mdpi.com/2226-4310/8/10/283 (visited on 01/26/2022).

[70] Matton, N., Paubel, P.-V., Puma, S. “Toward the Use of Pupillary Responses for Pilot Selection”. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 64.3 (May 2022), pp. 555–567. ISSN: 0018-7208, 1547-8181. DOI: 10.1177/0018720820945163. (Visited on 03/06/2024).

[71] Zheng, Z., Gao, S., Su, Y., Chen, Y., Wang, X. “Cognitive load-induced pupil dilation reflects potential flight ability”. In: *Current Psychology* (Aug. 12, 2022). ISSN: 1046-1310, 1936-4733. DOI: 10.1007/

s12144-022-03430-2. URL: https://link.springer.com/10.1007/s12144-022-03430-2 (visited on 11/30/2022).

[72] Lutnyk, L., Rudi, D., Schinazi, V. R., Kiefer, P., Raubal, M. "The effect of flight phase on electrodermal activity and gaze behavior: A simulator study". In: *Applied Ergonomics* 109 (May 2023), p. 103989. ISSN: 00036870. DOI: 10.1016/j.apergo.2023.103989. URL: https://linkinghub.elsevier.com/retrieve/pii/S0003687023000273 (visited on 03/11/2024).

[73] Chen, H., Liu, S., Wanyan, X., Pang, L., Dang, Y., Zhu, K., Yu, X. "Influencing factors of novice pilot SA based on DEMATEL-AISM method: From pilots' view". In: *Heliyon* 9.2 (Feb. 2023), e13425. ISSN: 24058440. DOI: 10.1016/j.heliyon.2023.e13425. URL: https://linkinghub.elsevier.com/retrieve/pii/S2405844023006321 (visited on 03/28/2024).

[74] Li, W.-C., Zhang, J., Court, S., Kearney, P., Braithwaite, G. "The influence of augmented reality interaction design on Pilot's perceived workload and situation awareness". In: *International Journal of Industrial Ergonomics* 92 (Nov. 2022), p. 103382. ISSN: 01698141. DOI: 10.1016/j.ergon.2022.103382. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169814122001238 (visited on 06/12/2024).

[75] Luo, R., Weng, Y., Jayakumar, P., Brudnak, M. J., Paul, V., Desaraju, V. R., Stein, J. L., Ersal, T., Yang, X. J. "Real-Time Workload Estimation Using Eye Tracking: A Bayesian Inference Approach". In: *International Journal of Human–Computer Interaction* (May 4, 2023), pp. 1–

16. ISSN: 1044-7318, 1532-7590. DOI: 10.1080/10447318.2023.2205274. URL: https://www.tandfonline.com/doi/full/10.1080/10447318.2023.2205274 (visited on 03/28/2024).

[76] Zhang, X., Qu, X., Xue, H., Tao, D., Li, T. "Effects of time of day and taxi route complexity on navigation errors: An experimental study". In: *Accident Analysis & Prevention* 125 (Apr. 2019), pp. 14–19. ISSN: 00014575. DOI: 10.1016/j.aap.2019.01.019. URL: https://linkinghub.elsevier.com/retrieve/pii/S000145751930123X (visited on 06/16/2022).

[77] Li, F., Chen, C.-H., Zheng, P., Feng, S., Xu, G., Khoo, L. P. "An explorative context-aware machine learning approach to reducing human fatigue risk of traffic control operators". In: *Safety Science* 125 (May 2020), p. 104655. ISSN: 09257535. DOI: 10.1016/j.ssci.2020.104655. URL: https://linkinghub.elsevier.com/retrieve/pii/S0925753520300527 (visited on 12/14/2021).

[78] Linardatos, P., Papastefanopoulos, V., Kotsiantis, S. "Explainable AI: A Review of Machine Learning Interpretability Methods". In: *Entropy* 23.1 (Dec. 25, 2020), p. 18. ISSN: 1099-4300. DOI: 10.3390/e23010018. URL: https://www.mdpi.com/1099-4300/23/1/18 (visited on 02/13/2024).

[79] LeCun, Y., Bengio, Y., Hinton, G. "Deep learning". In: *Nature* 521.7553 (May 28, 2015), pp. 436–444. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature14539. URL: https://www.nature.com/articles/nature14539 (visited on 02/13/2024).

[80] Polikar, R. “Ensemble Learning”. In: *Ensemble Machine Learning*. Ed. by Cha Zhang and Yunqian Ma. New York, NY: Springer New York, 2012, pp. 1–34. ISBN: 978-1-4419-9325-0 978-1-4419-9326-7. DOI: 10.1007/978-1-4419-9326-7_1. URL: https://link.springer.com/10.1007/978-1-4419-9326-7_1 (visited on 02/13/2024).

[81] Di Stasi, L., Marchitto, M., Antolí, A., Cañas, J. “Saccadic peak velocity as an alternative index of operator attention: A short review”. In: *European Review of Applied Psychology* 63.6 (Nov. 2013), pp. 335–343. ISSN: 11629088. DOI: 10.1016/j.erap.2013.09.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1162908813000741 (visited on 01/26/2023).

[82] Di Flumeri, G., De Crescenzio, F., Berberian, B., Ohneiser, O., Kramer, J., Aricò, P., Borghini, G., Babiloni, F., Bagassi, S., Piastra, S. “Brain–Computer Interface-Based Adaptive Automation to Prevent Out-Of-The-Loop Phenomenon in Air Traffic Controllers Dealing With Highly Automated Systems”. In: *Frontiers in Human Neuroscience* 13 (Sept. 6, 2019), p. 296. ISSN: 1662-5161. DOI: 10.3389/fnhum.2019.00296. URL: https://www.frontiersin.org/article/10.3389/fnhum.2019.00296/full (visited on 01/12/2023).

[83] Chouchane, H., Nakamura, H., Sato, K., Antona-Makoshi, J., Abe, G., Itoh, M. “Identifying the Out Of The Loop phenomenon during driving automation using spontaneous gaze behavior”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 66.1 (Sept. 2022), pp. 305–309. ISSN: 2169-5067, 1071-1813. DOI: 10.1177/1071181322661178.

URL: http://journals.sagepub.com/doi/10.1177/1071181322661178 (visited on 07/20/2023).

[84] Liu, X., Zheng, Y., Du, Z., Ding, M., Qian, Y., Yang, Z., Tang, J. "GPT understands, too". In: *AI Open* (Aug. 2023), S2666651023000141. ISSN: 26666510. DOI: 10.1016/j.aiopen.2023.08.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S2666651023000141 (visited on 09/09/2024).

[85] Wang, Y., Lipka, N., Rossi, R. A., Siu, A., Zhang, R., Derr, T. "Knowledge Graph Prompting for Multi-Document Question Answering". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 38.17 (Mar. 24, 2024), pp. 19206–19214. ISSN: 2374-3468, 2159-5399. DOI: 10.1609/aaai.v38i17.29889. URL: https://ojs.aaai.org/index.php/AAAI/article/view/29889 (visited on 09/09/2024).

[86] Lemieux, C., Park, D., Beschastnikh, I. "General LTL Specification Mining (T)". In: *2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. 2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE). Lincoln, NE, USA: IEEE, Nov. 2015, pp. 81–92. ISBN: 978-1-5090-0025-8. DOI: 10.1109/ASE.2015.71. URL: http://ieeexplore.ieee.org/document/7371998/ (visited on 11/27/2023).

[87] Kasenberg, D., Scheutz, M. "Interpretable apprenticeship learning with temporal logic specifications". In: *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. 2017 IEEE 56th Annual Conference on Decision and Control (CDC). Melbourne, Australia: IEEE, Dec. 2017,

pp. 4914–4921. ISBN: 978-1-5090-2873-3. DOI: 10.1109/CDC.2017.8264386. URL: http://ieeexplore.ieee.org/document/8264386/ (visited on 11/27/2023).

[88] Gabel, M., Su, Z. "Javert: fully automatic mining of general temporal properties from dynamic traces". In: *Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of software engineering*. SIGSOFT '08/FSE-16: SIGSOFT 2008 -16th International Symposium on the Foundations of Software Engineering. Atlanta Georgia: ACM, Nov. 9, 2008, pp. 339–349. ISBN: 978-1-59593-995-1. DOI: 10.1145/1453101.1453150. URL: https://dl.acm.org/doi/10.1145/1453101.1453150 (visited on 11/30/2023).

[89] Lo, D., Khoo, S.-C., Liu, C. "Mining temporal rules for software maintenance". In: *Journal of Software Maintenance and Evolution: Research and Practice* 20.4 (July 2008), pp. 227–247. ISSN: 1532-060X, 1532-0618. DOI: 10.1002/smr.375. URL: https://onlinelibrary.wiley.com/doi/10.1002/smr.375 (visited on 11/30/2023).

[90] Shi, J., Xiong, J., Huang, Y. "General past-time linear temporal logic specification mining". In: *CCF Transactions on High Performance Computing* 3.4 (Dec. 2021), pp. 393–406. ISSN: 2524-4922, 2524-4930. DOI: 10.1007/s42514-021-00079-4. URL: https://link.springer.com/10.1007/s42514-021-00079-4 (visited on 11/27/2023).

[91] Luo, W., Liang, P., Du, J., Wan, H., Peng, B., Zhang, D. "Bridging LTLf Inference to GNN Inference for Learning LTLf Formulae". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 36.9 (June 28,

2022), pp. 9849–9857. ISSN: 2374-3468, 2159-5399. DOI: 10.1609/aaai.v36i9.21221. URL: https://ojs.aaai.org/index.php/AAAI/article/view/21221 (visited on 10/07/2023).

[92] Zhang, Y., Li, F., Chang, D. "VR rehabilitation system evaluator: A fNIRS-based and LLM-enabled evaluation paradigm for Mild Cognitive Impairment". In: *Advanced Engineering Informatics* 62 (Oct. 2024), p. 102734. ISSN: 14740346. DOI: 10.1016/j.aei.2024.102734. URL: https://linkinghub.elsevier.com/retrieve/pii/S1474034624003823 (visited on 10/08/2024).

[93] Kim, J. W., Alaa, A., Bernardo, D. *EEG-GPT: Exploring Capabilities of Large Language Models for EEG Classification and Interpretation*. Version Number: 2. 2024. DOI: 10.48550/ARXIV.2401.18006. URL: https://arxiv.org/abs/2401.18006 (visited on 10/08/2024).

[94] Chen, H., Zeng, W., Chen, C., Cai, L., Wang, F., Wang, L., Zhang, W., Li, Y., Yan, H., Siok, W. T., Wang, N. *EEG Emotion Copilot: Pruning LLMs for Emotional EEG Interpretation with Assisted Medical Record Generation*. Version Number: 1. 2024. DOI: 10.48550/ARXIV.2410.00166. URL: https://arxiv.org/abs/2410.00166 (visited on 10/08/2024).

[95] Zhang, Y., Li, Q., Nahata, S., Jamal, T., Cheng, S.-K., Cauwenberghs, G., Jung, T.-P. "Integrating Large Language Model, EEG, and Eye-Tracking for Word-Level Neural State Classification in Reading Comprehension". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 32 (2024), pp. 3465–3475. ISSN: 1534-4320, 1558-0210.

DOI: 10.1109/TNSRE.2024.3435460. URL: https://ieeexplore.ieee.org/document/10636286/ (visited on 10/08/2024).

[96] Boiarshinov, D., Guajardo, J., Lanning, G. "Providing LLM-generated Point of Interest Description Based on Gaze Tracking". In: (2024).

[97] Fourati, S., Jaafar, W., Baccar, N., Alfattani, S. *XLM for Autonomous Driving Systems: A Comprehensive Review*. Sept. 16, 2024. arXiv: 2409.10484[cs,eess]. URL: http://arxiv.org/abs/2409.10484 (visited on 10/08/2024).

[98] Luo, R., Weng, Y., Wang, Y., Jayakumar, P., Brudnak, M. J., Paul, V., Desaraju, V. R., Stein, J. L., Ersal, T., Yang, X. J. "A workload adaptive haptic shared control scheme for semi-autonomous driving". In: *Accident Analysis & Prevention* 152 (Mar. 2021), p. 105968. ISSN: 00014575. DOI: 10.1016/j.aap.2020.105968. URL: https://linkinghub.elsevier.com/retrieve/pii/S0001457520317887 (visited on 03/28/2024).

[99] Wang, Y., Li, W.-C., Korek, W. T., Braithwaite, G. "Future flight deck design: Developing an innovative touchscreen inceptor combined with the primary flight display". In: *International Journal of Industrial Ergonomics* 101 (May 2024), p. 103588. ISSN: 01698141. DOI: 10.1016/j.ergon.2024.103588. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169814124000441 (visited on 06/12/2024).

[100] Wanyan, X., Zhuang, D., Lin, Y., Xiao, X., Song, J.-W. "Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation". In: *International Journal of Industrial*

Ergonomics 64 (Mar. 2018), pp. 1–7. ISSN: 01698141. DOI: 10.1016/j.ergon.2017.08.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169814117304201 (visited on 03/28/2024).

[101] Li, W.-C., Braithwaite, G., Wang, T., Yung, M., Kearney, P. "The benefits of integrated eye tracking with airborne image recorders in the flight deck: A rejected landing case study". In: *International Journal of Industrial Ergonomics* 78 (July 2020). ISSN: 0169-8141. DOI: 10.1016/j.ergon.2020.102982.

[102] LaBerge, D. L. "Attention". In: *Psychological Science* 1.3 (May 1990), pp. 156–162. ISSN: 0956-7976, 1467-9280. DOI: 10.1111/j.1467-9280.1990.tb00188.x. URL: http://journals.sagepub.com/doi/10.1111/j.1467-9280.1990.tb00188.x (visited on 03/09/2024).

[103] Wickens, C. D., Flach, J. M. "Information processing." In: *Human factors in aviation.* Academic Press series in cognition and perception. San Diego, CA, US: Academic Press, 1988, pp. 111–155. ISBN: 0-12-750030-8 (Hardcover).

[104] Caird, J. K., Simmons, S. M., Wiley, K., Johnston, K. A., Horrey, W. J. "Does Talking on a Cell Phone, With a Passenger, or Dialing Affect Driving Performance? An Updated Systematic Review and Meta-Analysis of Experimental Studies". In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 60.1 (Feb. 2018), pp. 101–133. ISSN: 0018-7208, 1547-8181. DOI: 10.1177/0018720817748145. (Visited on 03/09/2024).

[105] Horrey, W. J., Wickens, C. D. "Examining the Impact of Cell Phone Conversations on Driving Using Meta-Analytic Techniques". In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 48.1 (Mar. 2006), pp. 196–205. ISSN: 0018-7208, 1547-8181. DOI: 10.1518/001872006776412135. (Visited on 03/09/2024).

[106] Park, J., Park, W. "Functional requirements of automotive head-up displays: A systematic review of literature from 1994 to present". In: *Applied Ergonomics* 76 (Apr. 2019), pp. 130–146. ISSN: 00036870. DOI: 10.1016/j.apergo.2018.12.017. URL: https://linkinghub.elsevier.com/retrieve/pii/S0003687018307592 (visited on 03/28/2024).

[107] Lu, S. A., Wickens, C. D., Prinet, J. C., Hutchins, S. D., Sarter, N., Sebok, A. "Supporting Interruption Management and Multimodal Interface Design: Three Meta-Analyses of Task Performance as a Function of Interrupting Task Modality". In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 55.4 (Aug. 2013), pp. 697–724. ISSN: 0018-7208, 1547-8181. DOI: 10.1177/0018720813476298. (Visited on 03/10/2024).

[108] Ahmadi, N., Romoser, M., Salmon, C. "Improving the tactical scanning of student pilots: A gaze-based training intervention for transition from visual flight into instrument meteorological conditions". In: *Applied Ergonomics* 100 (Apr. 2022). ISSN: 0003-6870. DOI: 10.1016/j.apergo.2021.103642.

[109] Li, W.-C., Zhang, J., Minh, T. L., Cao, J., Wang, L. "Visual scan patterns reflect to human-computer interactions on processing different types

of messages in the flight deck". In: *International Journal of Industrial Ergonomics* 72 (July 2019), pp. 54–60. ISSN: 0169-8141. DOI: 10.1016/j.ergon.2019.04.003.

[110] Li, Q., Chen, C.-H., Ng, K. K., Yuan, X., Yiu, C. Y. "Single-pilot operations in commercial flight: Effects on neural activity and visual behaviour under abnormalities and emergencies". In: *Chinese Journal of Aeronautics* (Apr. 2024), S1000936124001377. ISSN: 10009361. DOI: 10.1016/j.cja.2024.04.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S1000936124001377 (visited on 06/12/2024).

[111] Okine, E. A., Zarei, E., Roggow, B. J. "Exploring the intellectual insights in aviation safety research: A systematic literature and bibliometric review". In: *Safety Science* 170 (Feb. 2024), p. 106354. ISSN: 09257535. DOI: 10.1016/j.ssci.2023.106354. URL: https://linkinghub.elsevier.com/retrieve/pii/S0925753523002965 (visited on 02/29/2024).

[112] Administration, F. A., Soucie, D. *Airplane Flying Handbook: FAA-H-8083-3A*. 1st ed. New York, NY: Skyhorse Publishing, 2014. ISBN: 978-1-62914-854-0.

[113] Bernhardt, K. A., Poltavski, D., Petros, T., Ferraro, F. R., Jorgenson, T., Carlson, C., Drechsel, P., Iseminger, C. "The effects of dynamic workload and experience on commercially available EEG cognitive state metrics in a high-fidelity air traffic control environment". In: *Applied Ergonomics* 77 (May 2019), pp. 83–91. ISSN: 00036870. DOI: 10.1016/j.apergo.2019.01.008. URL: https://linkinghub.elsevier.com/retrieve/pii/S000368701930016X (visited on 02/29/2024).

[114] Kearney, P., Li, W.-C., Yu, C.-S., Braithwaite, G. “The impact of alerting designs on air traffic controller’s eye movement patterns and situation awareness”. In: *Ergonomics* 62.2 (2019). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00140139.2018.1493151, pp. 305–318. DOI: 10.1080/00140139.2018.1493151. URL: https://doi.org/10.1080/00140139.2018.1493151.

[115] Wang, Y., Hu, R., Lin, S., Schultz, M., Delahaye, D. “The Impact of Automation on Air Traffic Controller’s Behaviors”. In: *Aerospace* 8.9 (Sept. 13, 2021), p. 260. ISSN: 2226-4310. DOI: 10.3390/aerospace8090260. URL: https://www.mdpi.com/2226-4310/8/9/260 (visited on 06/14/2024).

[116] **Lyu, M.**, Li, F., Lee, C.-H. “The Effects of Adaptive Automation on Pilots’ Flight Control Performance and Visual Attention Distribution”. In: *Advances in Transdisciplinary Engineering*. Ed. by Pisut Koomsap, Adam Cooper, and Josip Stjepandić. IOS Press, Nov. 7, 2023. ISBN: 978-1-64368-440-6 978-1-64368-441-3. DOI: 10.3233/ATDE230600. URL: https://ebooks.iospress.nl/doi/10.3233/ATDE230600 (visited on 11/24/2023).

[117] Huang, W., Xu, Y., Hildebrandt, M., Lau, N. “Comparing Eye-Gaze Metrics of Mental Workload in Monitoring Process Plants”. In: *Engineering Psychology and Cognitive Ergonomics*. Ed. by Don Harris. Vol. 11571. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 55–65. ISBN: 978-3-030-22506-3 978-3-030-22507-0. DOI: 10.1007/978-3-030-22507-0_5. URL: https:

//link.springer.com/10.1007/978-3-030-22507-0_5 (visited on 06/14/2024).

[118] Sturman, D., Wiggins, M. W., Auton, J. C., Loft, S. "Cue utilization differentiates resource allocation during sustained attention simulated rail control tasks." In: *Journal of Experimental Psychology: Applied* 25.3 (Sept. 2019), pp. 317–332. ISSN: 1939-2192, 1076-898X. DOI: 10.1037/xap0000204. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/xap0000204 (visited on 06/14/2024).

[119] Nakayama, M., Takahashi, K., Shimizu, Y. "The act of task difficulty and eye-movement frequency for the 'Oculo-motor indices'". In: *Proceedings of the symposium on Eye tracking research & applications - ETRA '02*. the symposium. New Orleans, Louisiana: ACM Press, 2002, p. 37. ISBN: 978-1-58113-467-4. DOI: 10.1145/507072.507080. URL: http://portal.acm.org/citation.cfm?doid=507072.507080 (visited on 06/13/2023).

[120] Holmqvist, K., Andersson, R. *Eye-tracking: A comprehensive guide to methods, paradigms and measures*. Oxford University Press, Sept. 22, 2011. ISBN: ISBN-13: 978-1979484893. URL: http://ukcatalogue.oup.com/product/9780199697083.do.

[121] Marchitto, M., Di Stasi, L. L., Cañas, J. J. "Ocular movements under taskload manipulations: Influence of geometry on saccades in air traffic control simulated tasks". In: *Human Factors and Ergonomics in Manufacturing & Service Industries* 22.5 (Sept. 2012), pp. 407–419. ISSN: 1090-8471, 1520-6564. DOI: 10.1002/hfm.20320. URL: https://

onlinelibrary.wiley.com/doi/10.1002/hfm.20320 (visited on 06/14/2024).

[122] Marchitto, M., Benedetto, S., Baccino, T., Cañas, J. J. "Air traffic control: Ocular metrics reflect cognitive complexity". In: *International Journal of Industrial Ergonomics* 54 (July 2016), pp. 120–130. ISSN: 01698141. DOI: 10.1016/j.ergon.2016.05.010. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169814116300452 (visited on 12/02/2022).

[123] Pedrotti, M., Mirzaei, M. A., Tedesco, A., Chardonnet, J.-R., Mérienne, F., Benedetto, S., Baccino, T. "Automatic Stress Classification With Pupil Diameter Analysis". In: *International Journal of Human-Computer Interaction* 30.3 (Mar. 4, 2014), pp. 220–236. ISSN: 1044-7318, 1532-7590. DOI: 10.1080/10447318.2013.848320. URL: http://www.tandfonline.com/doi/abs/10.1080/10447318.2013.848320 (visited on 05/27/2023).

[124] Niu, Y., Li, Z., Fan, Y. "Analysis of truck drivers' unsafe driving behaviors using four machine learning methods". In: *International Journal of Industrial Ergonomics* 86 (Nov. 2021), p. 103192. ISSN: 01698141. DOI: 10.1016/j.ergon.2021.103192. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169814121001104 (visited on 03/20/2024).

[125] Jiao, Y., Wang, X., Kang, Y., Zhong, Z., Chen, W. "A quick identification model for assessing human anxiety and thermal comfort based on physiological signals in a hot and humid working environment". In: *International Journal of Industrial Ergonomics* 94 (Mar. 2023), p. 103423.

ISSN: 01698141. DOI: 10.1016/j.ergon.2023.103423. URL: https://linkinghub.elsevier.com/retrieve/pii/S016981412300015X (visited on 03/20/2024).

[126] Lundberg, S., Lee, S.-I. *A Unified Approach to Interpreting Model Predictions*. Nov. 24, 2017. arXiv: 1705.07874[cs, stat]. URL: http://arxiv.org/abs/1705.07874 (visited on 10/04/2023).

[127] Collinson, R. P. G. "Autopilots and flight management systems". In: *Introduction to Avionics*. Ed. by R. P. G. Collinson. Dordrecht: Springer Netherlands, 1996, pp. 366–405. ISBN: 978-94-010-4007-5 978-94-011-0007-6. DOI: 10.1007/978-94-011-0007-6_8. URL: http://link.springer.com/10.1007/978-94-011-0007-6_8 (visited on 01/30/2023).

[128] Arico, P., Borghini, G., Di Flumeri, G., Bonelli, S., Golfetti, A., Graziani, I., Pozzi, S., Imbert, J.-P., Granger, G., Benhacene, R., Schaefer, D., Babiloni, F. "Human Factors and Neurophysiological Metrics in Air Traffic Control: A Critical Review". In: *IEEE Reviews in Biomedical Engineering* 10 (2017), pp. 250–263. ISSN: 1937-3333, 1941-1189. DOI: 10.1109/RBME.2017.2694142. URL: http://ieeexplore.ieee.org/document/7898410/ (visited on 01/30/2023).

[129] Endsley, M. R., Kaber, D. B. "Level of automation effects on performance, situation awareness and workload in a dynamic control task". In: *Ergonomics* 42.3 (Mar. 1, 1999), pp. 462–492. ISSN: 0014-0139, 1366-5847. DOI: 10.1080/001401399185595. URL: https://www.tandfonline.com/doi/full/10.1080/001401399185595 (visited on 01/30/2023).

[130] Wickens, C. D. “Cognitive factors in aviation.” In: *Handbook of applied cognition.* New York, NY, US: John Wiley & Sons Ltd, 1999, pp. 247–282. ISBN: 0-471-97765-9 (Hardcover).

[131] Sarter, N. B., Woods, D. D. “How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control”. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37.1 (Mar. 1995), pp. 5–19. ISSN: 0018-7208, 1547-8181. DOI: `10.1518/001872095779049516`. URL: `http://journals.sagepub.com/doi/10.1518/001872095779049516` (visited on 11/19/2023).

[132] Amalberti, R. “Automation in aviation: A human factors perspective”. In: *Handbook of aviation human factors* (1999). Publisher: Lawrence Erlbaum Associates Mahwah, NJ, pp. 173–192.

[133] Norman, D. A. “The Human Side of Automation”. In: *Road Vehicle Automation 2*. Ed. by Gereon Meyer and Sven Beiker. Series Title: Lecture Notes in Mobility. Cham: Springer International Publishing, 2015, pp. 73–79. ISBN: 978-3-319-19077-8 978-3-319-19078-5. DOI: `10.1007/978-3-319-19078-5_7`. URL: `https://link.springer.com/10.1007/978-3-319-19078-5_7` (visited on 11/20/2023).

[134] Berberian, B., Somon, B., Sahaï, A., Gouraud, J. “The out-of-the-loop Brain: A neuroergonomic approach of the human automation interaction”. In: *Annual Reviews in Control* 44 (2017), pp. 303–315. ISSN: 13675788. DOI: `10.1016/j.arcontrol.2017.09.010`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S1367578817301116` (visited on 11/20/2023).

[135] Li, F., Lee, C.-H., Chen, C.-H., Khoo, L. P. "Hybrid data-driven vigilance model in traffic control center using eye-tracking data and context data". In: *Advanced Engineering Informatics* 42 (Oct. 2019), p. 100940. ISSN: 14740346. DOI: 10.1016/j.aei.2019.100940. URL: https://linkinghub.elsevier.com/retrieve/pii/S1474034619300540 (visited on 12/14/2021).

[136] Mittner, M., Boekel, W., Tucker, A. M., Turner, B. M., Heathcote, A., Forstmann, B. U. "When the Brain Takes a Break: A Model-Based Analysis of Mind Wandering". In: *The Journal of Neuroscience* 34.49 (Dec. 3, 2014), pp. 16286–16295. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.2062-14.2014. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.2062-14.2014 (visited on 11/08/2023).

[137] Bixler, R., D'Mello, S. "Automatic gaze-based user-independent detection of mind wandering during computerized reading". In: *User Modeling and User-Adapted Interaction* 26.1 (Mar. 2016), pp. 33–68. ISSN: 0924-1868, 1573-1391. DOI: 10.1007/s11257-015-9167-1. URL: http://link.springer.com/10.1007/s11257-015-9167-1 (visited on 11/08/2023).

[138] Zheng, X., Yu, C., Zhang, M. "Lifelong reinforcement learning with temporal logic formulas and reward machines". In: *Knowledge-Based Systems* 257 (Dec. 2022), p. 109650. ISSN: 09507051. DOI: 10.1016/j.knosys.2022.109650. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950705122008358 (visited on 02/21/2024).

[139] Kim, J., Muise, C., Shah, A., Agarwal, S., Shah, J. “Bayesian Inference of Linear Temporal Logic Specifications for Contrastive Explanations”. In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. Twenty-Eighth International Joint Conference on Artificial Intelligence {IJCAI-19}. Macao, China: International Joint Conferences on Artificial Intelligence Organization, Aug. 2019, pp. 5591–5598. ISBN: 978-0-9992411-4-1. DOI: 10.24963/ijcai.2019/776. URL: https://www.ijcai.org/proceedings/2019/776 (visited on 11/09/2023).

[140] Gaglione, J.-R., Neider, D., Roy, R., Topcu, U., Xu, Z. “Learning Linear Temporal Properties from Noisy Data: A MaxSAT-Based Approach”. In: *Automated Technology for Verification and Analysis*. Ed. by Zhe Hou and Vijay Ganesh. Vol. 12971. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, pp. 74–90. ISBN: 978-3-030-88884-8 978-3-030-88885-5. DOI: 10.1007/978-3-030-88885-5_6. URL: https://link.springer.com/10.1007/978-3-030-88885-5_6 (visited on 11/27/2023).

[141] Neider, D., Gavran, I. “Learning Linear Temporal Properties”. In: *2018 Formal Methods in Computer Aided Design (FMCAD)*. 2018 Formal Methods in Computer Aided Design (FMCAD). Austin, TX: IEEE, Oct. 2018, pp. 1–10. ISBN: 978-0-9835678-8-2. DOI: 10.23919/FMCAD.2018.8603016. URL: https://ieeexplore.ieee.org/document/8603016/ (visited on 11/27/2023).

[142] Shah, A., Kamath, P., Shah, J. A., Li, S. "Bayesian Inference of Temporal Task Specifications from Demonstrations". In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Vol. 31. Curran Associates, Inc., 2018. URL: `https://proceedings.neurips.cc/paper_files/paper/2018/file/13168e6a2e6c84b4b7de9390c0ef5ec5-Paper.pdf`.

[143] Hessels, R. S., Kemner, C., Van Den Boomen, C., Hooge, I. T. C. "The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli". In: *Behavior Research Methods* 48.4 (Dec. 2016), pp. 1694–1712. ISSN: 1554-3528. DOI: `10.3758/s13428-015-0676-y`. URL: `http://link.springer.com/10.3758/s13428-015-0676-y` (visited on 11/26/2023).

[144] Li, Z., Li, Z., Li, F. "Visual Attention Analytics for Individual Perception Differences and Task Load-Induced Inattentional Blindness". In: *Cross-Cultural Design*. Ed. by Pei-Luen Patrick Rau. Vol. 14023. Series Title: Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2023, pp. 71–83. ISBN: 978-3-031-35938-5 978-3-031-35939-2. DOI: `10.1007/978-3-031-35939-2_6`. URL: `https://link.springer.com/10.1007/978-3-031-35939-2_6` (visited on 11/29/2023).

[145] Li, W.-C., Braithwaite, G., Yu, C.-S. "The Investigation of Pilots' Eye Scan Patterns on the Flight Deck during an Air-to-Surface Task". In: *Engineering Psychology and Cognitive Ergonomics, EPCE 2014*. Ed. by D

Harris. Vol. 8532. Lecture Notes in Computer Science. ISSN: 0302-9743. 2014, pp. 325–334. ISBN: 978-3-319-07515-0 978-3-319-07514-3.

[146] Haslbeck, A., Zhang, B. “I spy with my little eye: Analysis of airline pilots’ gaze patterns in a manual instrument flight scenario”. In: *Applied Ergonomics* 63 (Sept. 2017), pp. 62–71. ISSN: 00036870. DOI: 10.1016/j.apergo.2017.03.015. URL: https://linkinghub.elsevier.com/retrieve/pii/S000368701730073X (visited on 06/17/2022).

[147] Jayawardena, G., Jayarathna, S. “Automated Filtering of Eye Movements Using Dynamic AOI in Multiple Granularity Levels:” in: *International Journal of Multimedia Data Engineering and Management* 12.1 (Jan. 2021), pp. 49–64. ISSN: 1947-8534, 1947-8542. DOI: 10.4018/IJMDEM.2021010104. URL: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJMDEM.2021010104 (visited on 11/29/2023).

[148] McClung, S. N., Kang, Z. “Characterization of Visual Scanning Patterns in Air Traffic Control”. In: *Computational Intelligence and Neuroscience* 2016 (2016), pp. 1–17. ISSN: 1687-5265, 1687-5273. DOI: 10.1155/2016/8343842. URL: http://www.hindawi.com/journals/cin/2016/8343842/ (visited on 06/18/2022).

[149] Peysakhovich, V., Ledegang, W., Houben, M., Groen, E. “Classification of flight phases based on pilots’ visual scanning strategies”. In: *2022 Symposium on Eye Tracking Research and Applications*. ETRA ’22. New York, NY, USA: Association for Computing Machinery, June 8, 2022, pp. 1–7. ISBN: 978-1-4503-9252-5. DOI: 10.1145/3517031.3529641.

URL: https://dl.acm.org/doi/10.1145/3517031.3529641 (visited on 05/23/2024).

[150] Kang, Z., Landry, S. J. "An Eye Movement Analysis Algorithm for a Multielement Target Tracking Task: Maximum Transition-Based Agglomerative Hierarchical Clustering". In: *IEEE Transactions on Human-Machine Systems* 45.1 (Feb. 2015), pp. 13–24. ISSN: 2168-2291, 2168-2305. DOI: 10.1109/THMS.2014.2363121. URL: http://ieeexplore.ieee.org/document/6960823/ (visited on 06/17/2022).

[151] Li, F., Katsumata, S., Lee, C.-H., Ye, Q., Dahana, W. D., Tu, R., Li, X. "Autoencoder-Enabled Potential Buyer Identification and Purchase Intention Model of Vacation Homes". In: *IEEE Access* 8 (2020), pp. 212383–212395. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.3037920. URL: https://ieeexplore.ieee.org/document/9258913/ (visited on 12/14/2021).

[152] Mohanty, A., Sussman, T. J. "Top-down modulation of attention by emotion". In: *Frontiers in Human Neuroscience* 7 (2013). ISSN: 1662-5161. DOI: 10.3389/fnhum.2013.00102. URL: http://journal.frontiersin.org/article/10.3389/fnhum.2013.00102/abstract (visited on 11/29/2023).

[153] Microsoft, *Microsoft Flight Simulator*. Jan. 1, 2024. URL: https://www.flightsimulator.com/.

[154] Tobii, *Tobii Pro Glasses 3*. Jan. 1, 2024. URL: https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3.

[155] Feldhütter, A., Gold, C., Schneider, S., Bengler, K. "How the Duration of Automated Driving Influences Take-Over Performance and Gaze Behavior". In: *Advances in Ergonomic Design of Systems, Products and Processes*. Ed. by Christopher Marc Schlick, Sönke Duckwitz, Frank Flemisch, Martin Frenz, Sinem Kuz, Alexander Mertens, and Susanne Mütze-Niewöhner. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017, pp. 309–318. ISBN: 978-3-662-53304-8 978-3-662-53305-5. DOI: 10.1007/978-3-662-53305-5_22. URL: http://link.springer.com/10.1007/978-3-662-53305-5_22 (visited on 01/15/2024).

[156] Tobii, *Tobii Pro Glasses 3 - Field Guide*. Jan. 1, 2024. URL: https://connect.tobii.com/s/field-guide-glasses3?language=en_US.

[157] Gontar, P., Homans, H., Rostalski, M., Behrend, J., Dehais, F., Bengler, K. "Are pilots prepared for a cyber-attack? A human factors approach to the experimental evaluation of pilots' behavior". In: *Journal of Air Transport Management* 69 (June 2018), pp. 26–37. ISSN: 09696997. DOI: 10.1016/j.jairtraman.2018.01.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0969699717300510 (visited on 06/17/2022).

[158] Kee, J., Knuth, M., Lahey, J. N., Palma, M. A. "Does eye-tracking have an effect on economic behavior?" In: *PLOS ONE* 16.8 (Aug. 5, 2021). Ed. by Pablo Brañas-Garza, e0254867. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0254867. URL: https://dx.plos.org/10.1371/journal.pone.0254867 (visited on 05/17/2024).

[159] Aust, J., Mitrovic, A., Pons, D. "Assessment of the Effect of Cleanliness on the Visual Inspection of Aircraft Engine Blades: An Eye Tracking Study". In: *Sensors* 21.18 (Sept. 2021). DOI: 10.3390/s21186135.

[160] Li, F., Chen, C.-H., Xu, G., Khoo, L. P., Liu, Y. "Proactive mental fatigue detection of traffic control operators using bagged trees and gaze-bin analysis". In: *Advanced Engineering Informatics* 42 (Oct. 2019), p. 100987. ISSN: 14740346. DOI: 10.1016/j.aei.2019.100987. URL: https://linkinghub.elsevier.com/retrieve/pii/S1474034619305609 (visited on 12/14/2021).

[161] Zhang, T., Huang, X., Zhao, W., Bian, S., Du, P. "LogPrompt: A Log-based Anomaly Detection Framework Using Prompts". In: *2023 International Joint Conference on Neural Networks (IJCNN)*. 2023 International Joint Conference on Neural Networks (IJCNN). Gold Coast, Australia: IEEE, June 18, 2023, pp. 1–8. ISBN: 978-1-66548-867-9. DOI: 10.1109/IJCNN54540.2023.10191948. URL: https://ieeexplore.ieee.org/document/10191948/ (visited on 09/16/2024).

[162] Lee, Y., Kim, J., Kang, P. "LAnoBERT: System log anomaly detection based on BERT masked language model". In: *Applied Soft Computing* 146 (Oct. 2023), p. 110689. ISSN: 15684946. DOI: 10.1016/j.asoc.2023.110689. URL: https://linkinghub.elsevier.com/retrieve/pii/S156849462300707X (visited on 09/16/2024).

[163] Tang, X., Zheng, Z., Li, J., Meng, F., Zhu, S.-C., Liang, Y., Zhang, M. *Large Language Models are In-Context Semantic Reasoners rather than Symbolic Reasoners*. Version Number: 2. 2023. DOI: 10.48550/ARXIV.

2305.14825. URL: https://arxiv.org/abs/2305.14825 (visited on 10/03/2024).

[164] Liu, B., Zhao, Q.-C., Ren, Y.-Y., Wang, Q.-J., Zheng, X.-L. "An elaborate algorithm for automatic processing of eye movement data and identifying fixations in eye-tracking experiments". In: *Advances in Mechanical Engineering* 10.5 (May 2018), p. 168781401877367. ISSN: 1687-8140, 1687-8140. DOI: 10.1177/1687814018773678. URL: http://journals.sagepub.com/doi/10.1177/1687814018773678 (visited on 09/16/2024).

[165] Ghose, U., Srinivasan, A. A., Boyce, W. P., Xu, H., Chng, E. S. "PyTrack: An end-to-end analysis toolkit for eye tracking". In: *Behavior Research Methods* 52.6 (Dec. 2020), pp. 2588–2603. ISSN: 1554-3528. DOI: 10.3758/s13428-020-01392-6. URL: https://link.springer.com/10.3758/s13428-020-01392-6 (visited on 09/16/2024).

[166] Su, J., Jiang, C., Jin, X., Qiao, Y., Xiao, T., Ma, H., Wei, R., Jing, Z., Xu, J., Lin, J. *Large Language Models for Forecasting and Anomaly Detection: A Systematic Literature Review*. Version Number: 1. 2024. DOI: 10.48550/ARXIV.2402.10350. URL: https://arxiv.org/abs/2402.10350 (visited on 09/09/2024).

[167] Van Der Lee, C., Gatt, A., Van Miltenburg, E., Krahmer, E. "Human evaluation of automatically generated text: Current trends and best practice guidelines". In: *Computer Speech & Language* 67 (May 2021), p. 101151. ISSN: 08852308. DOI: 10.1016/j.csl.2020.101151. URL: https:

//linkinghub.elsevier.com/retrieve/pii/S088523082030084X (visited on 10/20/2024).

[168] Krejtz, K., Szmidt, T., Duchowski, A. T., Krejtz, I. "Entropy-based statistical analysis of eye movement transitions". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA '14: Eye Tracking Research and Applications. Safety Harbor Florida: ACM, Mar. 26, 2014, pp. 159–166. ISBN: 978-1-4503-2751-0. DOI: 10.1145/2578153.2578176. URL: https://dl.acm.org/doi/10.1145/2578153.2578176 (visited on 10/23/2024).

[169] Krejtz, K., Duchowski, A., Krejtz, I., Szarkowska, A., Kopacz, A. "Discerning Ambient/Focal Attention with Coefficient *K*". In: *ACM Trans. Appl. Percept.* 13.3 (May 2016). Place: New York, NY, USA Publisher: Association for Computing Machinery. ISSN: 1544-3558. DOI: 10.1145/2896452. URL: https://doi.org/10.1145/2896452.

[170] Burch, M., Kull, A., Weiskopf, D. "AOI Rivers for Visualizing Dynamic Eye Gaze Frequencies". In: *Computer Graphics Forum* 32.3 (June 2013), pp. 281–290. ISSN: 0167-7055, 1467-8659. DOI: 10.1111/cgf.12115. URL: https://onlinelibrary.wiley.com/doi/10.1111/cgf.12115 (visited on 10/23/2024).

[171] Verdonck, T., Baesens, B., Óskarsdóttir, M., Vanden Broucke, S. "Special issue on feature engineering editorial". In: *Machine Learning* 113.7 (July 2024), pp. 3917–3928. ISSN: 0885-6125, 1573-0565. DOI: 10.1007/s10994-021-06042-2. URL: https://link.springer.com/10.1007/s10994-021-06042-2 (visited on 10/23/2024).

Appendices

Appendix I. LLM insturctions

You are an experienced flight instructor and ergonomist. You are experienced in identifying the pilot's status through his gaze patterns. Now I'll give you the eye-tracking data of a pilot in CSV format, you need to estimate whether the pilot is conducting normal monitoring (Neg) or abnormality detection (Pos) behaviors based on the knowledge below:

Each CSV file contains a 9*9 matrix, representing the pilot's visual behaviors in 30 seconds. The first 8 rows and columns represent different Areas Of Interest(AOI), as introduced below.

ECAM and MFD: ECAM and MFD: Electronic Centralized Aircraft Monitor (ECAM) and Multi-Function Display (MFD), are used to monitor aircraft systems and provide alerts to pilots about the aircraft's health and status. They serve multiple purposes, such as displaying maps, flight plans, weather, and aircraft systems information.

External: refers to the external view outside the cockpit that provides additional visual references to the pilots through the window.

FCU: Flight Control Unit, the interface that allows pilots to enter and control the autopilot system, setting and adjusting things like altitude, heading, and speed.

LDG: Landing Gear Panel, consists of the controls and indicators for the aircraft's landing gear system, displaying the status of the gear and any related warnings.

MCDU: Multifunction Control Display Unit, the interface for the Flight Management System (FMS), where pilots can input flight plans, perform calculations, and manage the aircraft's performance.

ND: Navigation Display, shows the aircraft's route in relation to navigation aids, the flight plan, and other important details such as weather radar.

PFD: Primary Flight Display, displays critical flight information such as speed, altitude, and attitude, along with navigational information.

Elsewhere: The other parts inside the cockpit that are not defined as specific AOIs.

The cell (m,n) represents the gaze fixation moved from the m-th AOI to the n-th AOI. For example, cell (2,1) represents the total durations of the fixations on "ECAM and MFD", where the previous fixations are on "FCU".

The 9th row and the 9th column represent the sum of the previous columns and rows. More specifically, the 9th row represents the total fixation durations of these AOIs, and the 9th column represents the fixation durations moving out from these AOIs.

Based on the previous research, we have three findings:

To help you estimate the pilot's status, the two attached files <Pos_range_round_X.CSV> and <Neg_range_round_X.CSV> are attached for your reference. The two files contain the empirical range of these fixations. The ranges of some cells can be overlapped, then you need to decide which better fits the given data and give an estimation. For example, if the positive range is (0,100) and the negative range is (0,200), and the given data is 50, then it is more likely that the data belongs to the positive case (the smaller range).
The file <p_value_round_X.CSV> provides the paired t-test results of the 81 cells between positive and negative cases. This suggests that the cells where the p-value <0.05 have significant differences between the positive and negative cases, leading to higher importance for making estimation.

Please estimate whether the pilot is conducting normal monitoring (Neg) or abnormality detection (Pos) behaviours and briefly give some reasons. Meanwhile, please also estimate where the abnormality existed (a specific AOI) based on the distribution of the fixations,

```
if you estimate it is a positive case.

Please organize your answer in a JSON object containing the following
keys:
"AD": if you estimate the pilot is conducting abnormality detection,
return "Pos", else return "Neg";
"reason": a concise explanation that supports your estimation. Based
on the fixations paid on the AOIs, you may suggest which AOI(s) are
likely to have abnormalities. Please note: the AOI being paid more
fixations does not necessarily suggest an abnormality, since these
AOIs may naturally attract more fixations in normal monitoring activities.
The more different from the negative range, the more possibility
of having an abnormality;
"response": just return "NA" if you estimate the pilot is just conducting
normal monitoring behaviour ("Neg"). If you deem the pilot is conducting
abnormality detection ("Pos"), make a response to help him better
handle the abnormality using the tongue like a human co-pilot.  The
response should be specific based on the reason you gave. For example,
You may use the questions such as 'Do you need support in monitoring
the ECAM and MFD?' and 'Is there anything wrong on LDG?'. If there
are multiple AOIs showing the possibility of having abnormalities
based on the data analysis, only pick the most possible one or two
in your response.
An example of the answer in JSON format is given below:
{
```

```
  "AD": "Pos",
  "reason": "The fixation duration on ECAM and MFD is significantly high compared to both normal and abnormal ranges, suggesting an unusual pattern that is more indicative of abnormality detection. The high fixation on ECAM and MFD likely indicates the pilot is focusing on system monitoring.",
  "response": "Is there anything wrong? Do you need my support in monitoring the ECAM and MFD? "
}
Please only reply in the required JSON object, no other content is needed in your response.
```

Appendix II. Responses from LLM

Ground Truth	**Estim-ation**	**Response**
ECAM & MFD	ECAM & MFD	Is there anything wrong with the ECAM and MFD, or ND? Do you need support?
ECAM & MFD	ECAM & MFD	Are there any specific concerns with ECAM and MFD? Do you need assistance?
ECAM & MFD	ECAM & MFD	Is there an issue requiring focused attention across systems like ECAM and MFD, or navigational displays? Do you need assistance handling the alerts or monitoring?

ECAM & MFD	ECAM & MFD	Is there anything wrong with the ECAM and MFD? Do you need support?
ECAM & MFD	ECAM & MFD	Are there any issues with the ECAM and MFD? Do you need assistance with monitoring it?
ECAM & MFD	ECAM & MFD	Are there any specific concerns with ECAM and MFD? Do you need assistance?
ECAM & MFD	ECAM & MFD	Do you need assistance with the ECAM and MFD, or any alerts from it?
ECAM & MFD	ECAM & MFD	Is there any issue with the ECAM and MFD? Is it showing any discrepancies?
ECAM & MFD	ECAM & MFD	Are there issues with the ECAM and MFDs? Do you need assistance with monitoring?
ECAM & MFD	ECAM & MFD	Is there an issue requiring focused attention across systems like ECAM and MFD, or navigational displays? Do you need assistance handling the alerts or monitoring?
ECAM & MFD	ECAM & MFD	Are there any issues requiring attention across systems like ECAM and MFD? Can I assist with monitoring?
ECAM & MFD	ECAM & MFD	Is there anything wrong? Do you need my support in monitoring the ECAM and MFD?
ECAM & MFD	ECAM & MFD	Is there anything wrong? Do you need my support in monitoring the ECAM and MFD?
ECAM & MFD	ECAM & MFD	Do you need support with the ECAM and MFD, or ND systems? Are there any alerts or unusual readings?

ECAM & MFD	ECAM & MFD	Is there an issue with the ECAM and MFD? Do you need help with monitoring?
ECAM & MFD	ECAM & MFD	Are there any alerts or unusual readings in the ECAM and MFD that need attention? Do you need additional support?
ECAM & MFD	ECAM & MFD	Are there any concerns with systems like ECAM and MFD, or PFD that need focus? Do you need assistance with handling alerts or monitoring?
ECAM & MFD	ND	Are there any concerns with the ND section that need attention? Do you need assistance monitoring operations?
ECAM & MFD	Others	Are there changes or anomalies in the FCU or MCDU systems? Do you need assistance monitoring the navigation or control displays?
ECAM & MFD	Others	Do you need my support with the PFD? Are there issues with the PFD or FCU settings?
ECAM & MFD	ECAM & MFD	Is there anything wrong? Do you need my support in monitoring the ECAM and MFD?
ECAM & MFD	ECAM & MFD	Is there an issue requiring focused attention across systems like ECAM and MFD, or navigational displays? Do you need assistance handling the alerts or monitoring?
ECAM & MFD	ECAM & MFD	Is there anything wrong with the ECAM and MFD? Do you need support?
ECAM & MFD	ECAM & MFD	Is there anything wrong? Do you need my support in monitoring the ECAM and MFD?

ECAM & MFD	ECAM & MFD	Are there any issues with the ECAM and MFD? Do you need assistance with monitoring it?
ECAM & MFD	ECAM & MFD	Do you need assistance with ECAM and MFD, or other navigational displays? Are there any issues with external conditions impacting flight stability?
ECAM & MFD	ECAM & MFD	Are there any specific concerns with ECAM and MFD? Do you need assistance?
ECAM & MFD	ECAM & MFD	Do you need assistance with the ECAM and MFD, or any alerts from it?
ECAM & MFD	ECAM & MFD	Is there any issue with the ECAM and MFD? Is it showing any discrepancies?
ECAM & MFD	ECAM & MFD	Are there issues with the ECAM and MFDs? Do you need assistance with monitoring?
LDG	LDG	Is everything alright with the LDG? Do you need help with monitoring or checking it?
LDG	LDG	Are there any issues with the LDG? Do you need support monitoring external conditions or landing gear status?
LDG	LDG	Is there anything wrong? Do you need my support in monitoring the LDG?
LDG	Others	Are there any specific concerns with MCDU? Do you need assistance?
LDG	LDG	Is everything alright with the LDG? Do you need help with monitoring or checking it?

LDG	ECAM & MFD	Is there an issue with the ECAM and MFD? Do you need help with monitoring?
LDG	ECAM & MFD	Is there an issue requiring focused attention across systems like ECAM and MFD, or navigational displays? Do you need assistance handling the alerts or monitoring?
LDG	LDG	Are there any concerns with LDG that need attention? Do you need assistance monitoring?
LDG	LDG	Is there an issue requiring focused attention on LDG? Do you need assistance handling the alerts or monitoring?
LDG	LDG	Do you need assistance with the LDG, or any alerts from it?
LDG	LDG	Is there anything wrong? Do you need my support in monitoring the LDG?
LDG	ECAM & MFD	Are there any issues requiring attention across systems like ECAM and MFD? Can I assist with monitoring?
LDG	LDG	Is there anything wrong? Do you need my support in monitoring the LDG?
LDG	LDG	Is there anything wrong? Do you need my support in monitoring the LDG?
LDG	Others	Do you need support with the MCDU, or PFD? Are there any alerts or unusual readings?
LDG	Others	Is there an issue with the ECAM and MFD? Do you need help with monitoring?

LDG	LDG	Are there any alerts on the LDG that need attention? Do you need additional support?
LDG	LDG	Is there any concern with LDG that need focus? Do you need assistance with handling alerts or monitoring?
LDG	LDG	Is there any concern with the LDG that need attention? Do you need assistance monitoring operations?
LDG	LDG	Are there changes or anomalies in the FCU or MCDU systems? Do you need assistance monitoring the navigation or control displays?
LDG	LDG	Are you experiencing any issues with the landing gear panel or ND? Do you require assistance in monitoring them?
LDG	LDG	Is there anything unusual with the LDG or ND? Could you provide more details about any abnormalities?
LDG	LDG	Is there anything wrong? Do you need my support in monitoring the LDG?
LDG	LDG	Do you need my support with the PFD? Are there issues with the PFD or FCU settings?
LDG	LDG	Is there anything wrong? Do you need my support in monitoring the LDG?
LDG	LDG	Is there anything wrong? Do you need my support in monitoring the LDG?
LDG	LDG	Do you need my support with the LDG? Are there issues with the LDG or FCU settings?

LDG	LDG	Do you need assistance with LDG? Are there any issues with external conditions impacting flight stability?
LDG	Others	Do you need support focusing on FCU or ND systems? Are there external conditions affecting navigation or displays that require attention?
ND	ND	Do you need my support with the ND? Are there issues with the ND?
ND	ND	Is there anything wrong? Do you need my support in monitoring the ND?
ND	ND	Is there an issue requiring focused attention across systems like ND, or PFD? Do you need assistance handling the alerts or monitoring?
ND	Others	Is there anything wrong on Elsewhere?
ND	ECAM & MFD	Is there anything wrong? Do you need my support in monitoring the ECAM and MFD?
ND	ND	Is there anything wrong? Do you need my support in monitoring the ND?
ND	ND	Is there anything wrong with the ND? Do you need support?
ND	ECAM & MFD	Do you need assistance with the ECAM and MFD, or any alerts from it?
ND	ND	Do you need assistance with ND?

ND	ND	Are there changes or anomalies in the ND or FCU systems? Do you need assistance monitoring the navigation or control displays?
ND	ND	Are there any issues with the ND? Do you need assistance with monitoring it?
ND	ND	Is there anything wrong? Do you need my support in monitoring the ND?
ND	ND	Are there any concerns with the ND section that need attention? Do you need assistance monitoring operations?
ND	ND	Are there any issues with the ND? Do you need assistance with monitoring it?
ND	ECAM & MFD	Are there any alerts or unusual readings in the ECAM and MFD that need attention? Do you need additional support?
ND	ECAM & MFD	Are there any concerns with systems like ECAM and MFD, or PFD that need focus? Do you need assistance with handling alerts or monitoring?
ND	ND	Do you need assistance with ND, or other navigational displays? Are there any issues with external conditions impacting flight stability?
ND	ND	Are there any specific concerns with ND? Do you need assistance?
ND	ND	Is there any issue with the ND? Is it showing any discrepancies?

ND	ND	Is there anything wrong? Do you need my support in monitoring the ND?
ND	ND	Is there anything wrong? Do you need my support in monitoring the ND or the PFD?
ND	ND	Is there anything wrong? Do you need my support in monitoring theND?
ND	Others	Are there any concerns with PFD that need attention? Do you need assistance monitoring?
ND	Others	Is there an issue requiring focused attention on MCDU? Do you need assistance handling the alerts or monitoring?
ND	ND	Are there issues with the ND? Do you need assistance with monitoring?
ND	ND	Is everything alright with the ND? Do you need help with monitoring or checking it?
ND	ND	Are there any issues with the LDG? Do you need support monitoring external conditions or landing gear status?
ND	ND	Is there anything wrong? Do you need my support in monitoring the LDG?
ND	ND	Are there any specific concerns with ND? Do you need assistance?
ND	ND	Is everything alright with the ND? Do you need help with monitoring or checking it?

ND	Others	Is there an issue with the ND? Do you need help with monitoring?
Normal	ECAM & MFD	Is there an issue with the ECAM and MFD? Do you need help with monitoring?
Normal	ND	Are there any issues with the ND? Do you need assistance with monitoring it?
Normal	ECAM & MFD	Is there an issue requiring focused attention across systems like ECAM and MFD, or navigational displays? Do you need assistance handling the alerts or monitoring?
Normal	ECAM & MFD	Do you need assistance with the ECAM and MFD, or any alerts from it?
Normal	ECAM & MFD	Is there any issue with the ECAM and MFD? Is it showing any discrepancies?
Normal	Others	Are there any specific concerns with MCDU? Do you need assistance?
Normal	Others	Is there an issue with the MCDU? Do you need help with monitoring?
Normal	ECAM & MFD	Are there issues with the ECAM and MFDs? Do you need assistance with monitoring?
Normal	Others	Do you need support focusing on FCU or ND systems? Are there external conditions affecting navigation or displays that require attention?

Normal	ND	Are there any concerns with the ND section that need attention? Do you need assistance monitoring operations?
Normal	Others	Are there changes or anomalies in the FCU or MCDU systems? Do you need assistance monitoring the navigation or control displays?
Normal	Others	Do you need my support with the PFD? Are there issues with the PFD or FCU settings?
Normal	ECAM & MFD	Is there an issue requiring focused attention across systems like ECAM and MFD, or navigational displays? Do you need assistance handling the alerts or monitoring?

Appendix III. Information sheet of experiment

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

INFORMATION SHEET

Psychophysiological Data-based Knowledge Transfer Model for Adaptive Training

You are invited to participate in the above project conducted by Dr Li Fan, who is a staff member of the Department of Aeronautical and Aviation Engineering in The Hong Kong Polytechnic University. The project has been approved by the Human Subjects Ethics Sub-committee (HSESC) of The Hong Kong Polytechnic University (HSESC Reference Number: HSEARS20211117002).

The aims/objectives of this project are to study the characteristics of biometric signals collected by EEG and Eye tracker; propose a novel approach (based on psychophysiological and behavioural data) to evaluate the situation awareness (SA) level of pilots and assess the sensitivity of this method; and to further reveal the relationship between SA level and biometric signals, providing a basis for real-time risk management in the future. The experiment may cover one or several areas, including:

- Situation awareness of response time, latency, mental fatigue and task performance

You are invited to take part in a procedure to investigate the situation awareness level of the subjects. Measurements will be taken by eye tracker and EEG equipment. The eye tracker measures the eyeball positions and movement. EEG measures the electrical activity in the brain (brain waves) using electrodes (small metal discs or sensors) placed on the head with gel. There should be minimal discomfort or risk. You will then be asked to complete a questionnaire, which will take you about 5 minutes. The test does not hurt and the whole investigation will take about 100 minutes.

The testing should not result in any undue discomfort, but you will need to perform the tasks required by the researchers. You are required to complete the tasks as the pilot role and perform flying activities in a flight simulator.

The information you provide as part of the project is the research data. Any research data from which you can be identified is known as personal data. Personal data does not include data where the identity has been removed (anonymous data). We will minimise our use of personal data in the study as much as possible. The researcher and his team, supervisor will have access to personal data and research data for the purposes of the study. Responsible members of The Hong Kong Polytechnic University may be given access for monitoring and/or audit of the research.

All information related to you will remain confidential. For data safety and confidentiality, the data will be stored in a locked computer. All the data, samples and study records will be stored with a password-protected zip. Only the investigator, the research staff and the research students under the investigator's supervision have the right to access the data in the period of conducting the research experiments and research project. All the data, samples and study records will be stored with a password-protected achiever. The information collected will be kept one year after project completion/publication or public release of the research results. The Hong Kong Polytechnic University takes reasonable precautions to prevent the loss, misappropriation, unauthorized access or destruction of the information you provide.

You have every right to withdraw from the study before or during the measurement without penalty of any kind.

If you have any questions, you may ask our helpers now or later, even after the study has started.

You may contact Dr Li Fan (tel. no.: 3400 2468/ email: fan-5.li@__________) of PolyU under the following situations:

a. if you have any other questions in relation to the study;
b. if, under very rare conditions, you become injured as a result of your participation in the study; or
c. if you want to get access to/or change your personal data before 31th Decemeber 2024.

In the event you have any complaints about the conduct of this research study, you may contact Miss Cherrie Mok, the Secretary of the Human Subjects Ethics Sub-Committee of The Hong Kong Polytechnic University by email (cherrie.mok@________) or in writing (c/o Research Office of The Hong Kong Polytechnic University) stating clearly the responsible person and department of this study as well as the HSESC Reference Number.

Thank you for your interest in participating in this study.

Dr Li Fan
Principal Investigator/Chief Investigator

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

參與者須知

用於自適應訓練的基於心理生理數據的知識轉移模型

我們誠邀閣下參與上述由香港理工大學航空及民航工程學系李凡博士開展研究。此研究已獲香港理工大學人類實驗對象操守小組委員會批准（人類實驗對象操守小組委員會項目參考編號: HSEARS20211117002）。

此研究的目的是研究腦電圖和眼動儀收集的生物特徵信號的特徵；提出一種新方法（基於心理生理和行為數據）來評估飛行員的態勢感知水平並評估該方法的敏感性； 並進一步揭示態勢感知水平與生物特徵信號之間的關係，為未來的實時風險管理提供依據。實驗涵蓋一個或多個領域，包括：

- 不同情景下的反應時間，延遲度，精神疲勞度和工作效能

過程中，研究人員將透過眼動儀及腦電圖等設備測量受試者的態勢感知程度。眼動儀用於測量眼球的位置和動態。腦電圖透過戴於頭上的電子儀器（小金屬圓盤或傳感器）測量大腦中的腦電波。受試者於過程中或感到輕微不適或有輕微風險，但實驗不會對受試者造成傷害。其後，受試者將需填寫一份問卷，大約需時 5 分鐘。整個實驗需時約 100 分鐘。

此實驗不會引致強烈不適，但受試者需根據研究人員的指引完成實驗。受試者將以空勤人員的身份模擬處理相應指令, 並於模擬器執飛指定任務。

實驗中收集的數據為「研究數據」。任何可以識別出受試者的數據為「個人數據」。個人數據並不包括已刪除身份識別資料的數據（ 即匿名數據）。我們將盡可能減少在研究中對個人數據的使用。只有項目負責人及其團隊可存取及瀏覽研究數據和個人數據進行研究。香港理工大學的相關人士亦會被授予監察和審核此研究項目的權限。

所有與受試者有關的資料將會被保密。為保障數據的安全性和保密性，所有數據將儲存在已鎖定的電腦設備裏。所有數據，樣本和研究記錄將額外使用密碼保存。在進行實驗和研究項目期間，只有相關研究人員可在項目負責人的監督下存取數據。我們所收集的數據將於項目完成或發布研究結果後保存一年。香港理工大學會採取合理的措施防止丟失、盜用、未經授權的存取或破壞所收集到的數據。

如您在實驗開始之前或實驗期間退出研究，您將不會受到任何形式的懲罰。

如您有任何疑問，可隨時向研究人員查詢。

若有以下情況，您可聯繫李凡博士（電話：34002468／電子郵件: fan-5.li@ ）：

a. 如果您有其他與此研究有關的問題；
b. 如果在非常罕見的情況下，您因參加此實驗而受傷；
c. 如果您想在 2024 年 12 月 31 日之前存取或更改您的個人數據。

如果您對此項研究有任何投訴，可透過電子郵件（cherrie.mok@ ）或郵件聯繫香港理工大學人類實驗對象操守小組委員會秘書莫小姐。若郵寄投訴，請於信封註明「轉交香港理工大學研究事務處」、研究負責人、及其所屬部門及人類實驗對象操守小組委員會項目參考編號。

我們再次感謝您的參與。

李凡博士
項目負責人

Hung Hom Kowloon Hong Kong 香港 九龍 紅磡
Tel 電話 (852) 2766 5111 Fax 傳真 (852) 2784 3374
Email 電郵 polyu@polyu.edu.hk
Website 網址 www.polyu.edu.hk

Appendix IV. Consent form of experiment

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

CONSENT TO PARTICIPATE IN RESEARCH

Psychophysiological Data-based Knowledge Transfer Model for Adaptive Training

I ______________________ hereby consent to participate in the captioned research conducted by Dr Li Fan.

I understand that information obtained from this research may be used in future research and published. However, my right to privacy will be retained, i.e. my personal details will not be revealed.

The procedure as set out in the attached information sheet has been fully explained. I understand the benefit and risks involved. My participation in the project is voluntary.

I acknowledge that I have the right to question any part of the procedure and can withdraw at any time without penalty of any kind.

Name of participant	______________________
Signature of participant	______________________
Name of Parent or Guardian (if applicable)	______________________
Signature of Parent or Guardian (if applicable)	______________________
Name of researcher	Dr. Li Fan
Signature of researcher	______________________
Date	______________________

Hung Hom Kowloon Hong Kong 香港 九龍 紅磡
Tel 電話 (852) 2766 5111 Fax 傳真 (852) 2784 3374
Email 電郵 polyu@polyu.edu.hk
Website 網址 www.polyu.edu.hk

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

參與研究同意書

用於自適應訓練的基於心理生理數據的知識轉移模型

本人____________________同意參與由李凡博士開展的上述研究。

本人知悉此研究所得的資料可能被用作日後的研究及發表，但本人的私隱權利將得以保留，即本人的個人資料不會被公開。

研究人員已向本人清楚解釋列在所附資料卡上的研究程序，本人明瞭當中涉及的利益及風險；本人自願參與研究項目。

本人知悉本人有權就程序的任何部分提出疑問，並有權隨時退出而不受任何懲處。

參與者姓名 ____
參與者簽署 ____
家長或監護人(如適用) 姓名 __________________________
家長或監護人(如適用) 簽署 __________________________
研究人員姓名 ______________________________________
研究人員簽署 ______________________________________
日期 ____

Hung Hom Kowloon Hong Kong 香港 九龍 紅磡
Tel 電話 (852) 2766 5111 Fax 傳真 (852) 2784 3374
Email 電郵 polyu@polyu.edu.hk
Website 網址 www.polyu.edu.hk