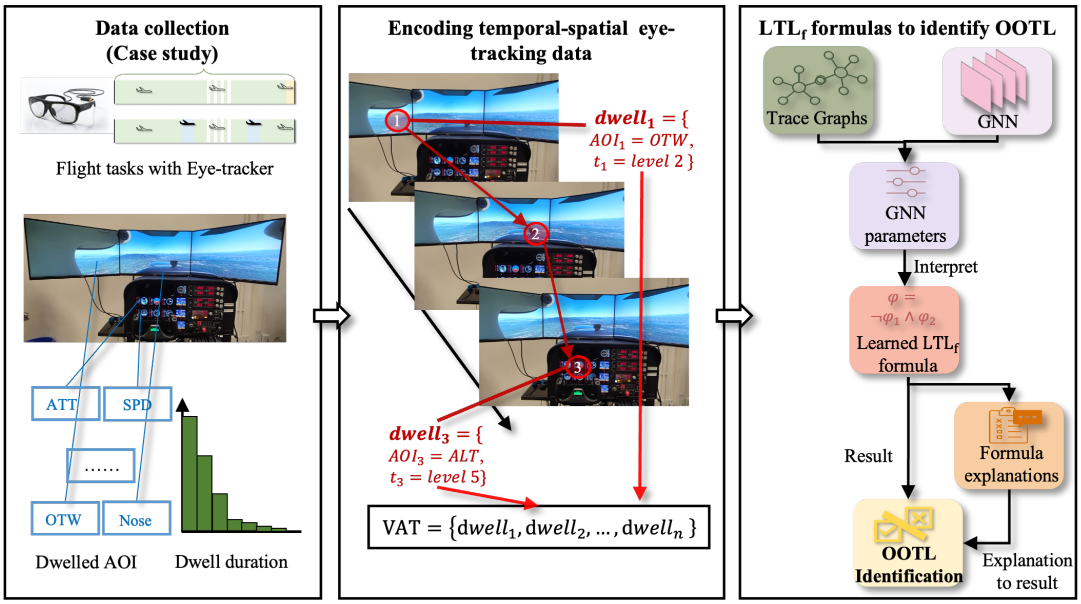
https://dx.doi.org/10.1016/j.knosys.2024.112086

Graphical Abstract

VALIO: Visual Attention-based Linear Temporal Logic method for Explainable Out-Of-The-Loop Identification

Mengtao Lyu, Fan Li, Ching-Hung Lee, Chun-Hsien Chen



VALIO: Visual Attention-based Linear Temporal Logic method for Explainable Out-Of-The-Loop Identification

Mengtao Lyu[a], Fan Li[a,*], Ching-Hung Lee[b], Chun-Hsien Chen[c]

[a]*Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region, 999077, China*
[b]*School of Public Policy and Administration, Xi'an Jiaotong University, Xi'an, China*
[c]*School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore*

Abstract

The phenomenon of being Out-Of-The-Loop (OOTL) can significantly undermine pilots' performance and pose a threat to aviation safety. Previous attempts to identify OOTL status have primarily utilized "black-box" machine learning techniques, which fail to provide explainable insights into their findings. To address this gap, our study introduces a novel application of Linear Temporal Logic (LTL) methods within a framework named Visual Attention LTL_f for Identifying OOTL (VALIO), leveraging eye-tracking technology to non-intrusively capture the pilots' attentional focus. By encoding Areas of Interest (AOIs) and gaze durations within the cockpit into Visual Attention Traces (VAT), the method captures the spatial and temporal dimensions of visual attention. It enables the LTL methods to generate interpretable formulas that classify pilot behaviors and provide insights into the understanding of the OOTL phenomenon. Through a case study of a simulated flight experiment, we compared the efficacy of this approach using different time windows from 10 seconds to 75 seconds. The results demonstrate that VALIO's performance is stable across all time windows with the best F1 score of 0.815 and the lowest F1 of 0.769. And it significantly outperforms the other machine learning methods when using time windows shorter than 30 seconds, signifying its ability to detect the OOTL status more in-timely. Moreover, the VALIO elucidates pilot behaviors through the derivation of human-readable

*Corresponding author
Email address: `fan-5.li@polyu.edu.hk` (Fan Li)

LTL_f formulas, offering the explainability of the results and insights into OOTL characteristics. Overall, this research proposes the VALIO framework as an improvement for OOTL identification in both performance and explainability.

Keywords: eye-tracking; flight safety; pilot performance; human-automation interaction

---

1. Introduction

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 [1]. 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 [2]. 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 [3, 4, 5]. 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 [6]. Data from both NASA and a survey among German aviators indicate that the OOTL phenomenon significantly contributes to human errors in aviation [7, 8].

While the problem of being OOTL has been acknowledged and extensively studied from diverse perspectives for several decades, there remains a significant gap in our understanding, particularly regarding the explainability of methods used to identify OOTL status [9, 10]. Traditional approaches often rely on machine learning models that interpret various cognitive states such as workload, fatigue, and situational awareness—using data collected from biometric sources like Electroencephalograms (EEG) and eye-tracking [11, 12]. However, the opaque nature of these "black-box" methodologies has raised concerns, especially within risk-sensitive fields such as aviation, where the inability to understand the reasoning behind decisions and the potential for errors could have severe consequences [13, 14]. This opacity 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 [15]. 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 study aimed to develop an OOTL identification approach with explainability based on eye-tracking methods.

This research innovatively integrates Linear Temporal Logic (LTL) and pilots' visual behaviours for the explainable identification of their OOTL status [16]. Investigating visual behaviours via eye-tracking technology offers a valuable window into pilots' attentional focus and cognitive status, which are key factors in detecting OOTL status [17, 18]. Unlike methods based on functional Near-Infrared Spectroscopy (fNIRS) and Electroencephalograms (EEG), which lack direct implications for behavioral 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). Moreover, eye-tracking technology facilitates non-invasive data collection, offering a significant advantage over other biometric approaches. Consequently, this study leverages eye-tracking to analyze the distribution of visual attention, aiming to accurately identify pilots' OOTL status with a focus on explainability and practical application [19].

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 [20, 21]. 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 [16].

Two primary challenges of achieving explainable OOTL identification based on eye-tracking data using LTL methods lay in the model compatibility and the performance with short lead time. The LTL method is designed to process data composed of atomic propositions, which indicate the occurrence of specific events. However, in real-world scenarios, pilots' gaze movements dynamically respond to external stimuli, and the data obtained from eye-tracking differs significantly from the proportion traces that the LTL method can handle. The raw eye-tracking data consists of continuous values (e.g., gaze location, pupil diameter, etc.) without clear boundaries to determine the existence of certain events. Even the processed eye-tracking data, which reflects the characteristics of visual behaviors over a certain period (e.g., average fixation duration, saccade velocity), is also composed of continuous values. Therefore, it is challenging to integrate eye-tracking data into the LTL methods for generating LTL formulas and verifying the pilots' status. Meanwhile, ensuring the performance with a short time window is another challenge for detection of the OOTL status in-time and more efficiently.

To tackle the compatibility challenge, 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 partitions the cockpit view into several Areas of Interest (AOIs) based on their functionalities and positions to evaluate where pilots have directed their gaze and the extent of attention they have allocated, considering both temporal and spatial aspects. To verify the effectiveness of the model in short time window and its stability across different time windows, a comparative flight simulation experiment is conducted as a case study. The performance of the proposed *Visual Attention* LTL_f *for Identifying OOTL (VALIO)* framework is compared with seven state-of-the-art machine learning methods. It is found that the VALIO method obtained stable performance across different time windows, and its superiority is affirmed by outperforming the other methods with shorter time windows.

The remainder of this paper is organized as follows: Section 2 reviews existing studies on the OOTL phenomenon, attention distribution studies based on eye-tracking technologies, and LTL methods. Section 3 details the framework of the proposed VALIO methods. Section 4 presents a case study to validate the proposed approach and discusses the results in comparison with other benchmark methods. Finally, Section 5 outlines the main contributions and limitations of this work and highlights future research directions.

2. Related work

2.1. *Out-Of-The-Loop phenomenon*

The aviation industry is heavily dependent on advanced automation systems to bolster efficiency, safety, and overall performance [1, 22]. However, the increasing complexity and usage of these systems have given rise to significant concerns about the Out-Of-The-Loop (OOTL) phenomenon [23, 24]. The OOTL phenomenon refers to situations where operators become disengaged from ongoing tasks, leading to potential performance degradation and safety risks [4, 3].

The OOTL phenomenon carries substantial implications, from reduced decision and action performance, to critical incidents and accidents [7, 8]. Operators experiencing the OOTL phenomenon may find it challenging to promptly identify system malfunctions or failures [25]. Such delayed or inappropriate responses to system anomalies can further intensify the severity of potential accidents [9]. Meanwhile, 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 [26, 27].

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 [28]. For example, vigilance failure has been identified as a critical factor of OOTL phenomenon [29, 30]. Additionally, mind-wandering (MW) has emerged as a significant aspect in the examination of OOTL, further expanding the scope of research in this area [9]. 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 [10]. Besides, the discussion has also been broadened with other related factors such as daydreaming and distraction, underscoring the complexity of OOTL phenomenon[31, 32]. 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 [33, 12, 34, 35, 36]. Among these, eye-tracking stands out as particularly promising for analyzing correlations and recognizing OOTL status, given its widespread application in addressing human-factor-related safety concerns within the aviation industry [11]. Its ability to assess pilots' visual attention on crucial flight information in a real-time, non-intrusive manner underscores its value in enhancing aviation safety.

Generally, previous works have studied the OOTL phenomenon 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 [37, 38]. However, there is an absence of explainability in the inference process these deep learning or ensemble models with complex hierarchical structures since the rationale behind their decisions are hard to understand and interpret[39, 40, 41]. Therefore, these "black-box" methods has sparked significant concerns about their application in the high-risk aviation industry [13]. Meanwhile, the OOTL identification results derived from purely digital numbers cannot provide direct guidance on how pilots should modify their behaviours to prevent or mitigate the OOTL phenomenon. To address this gap, it is crucial to capture the eye movement to assess the visual attention of the subjects, and generate more instructive recommendations for pilots on how they should manage their visual behaviours to counteract the OOTL phenomenon.

2.2. Linear Temporal Logic for behaviour classification

This study utilizes Linear Temporal Logic (LTL) to address the explainability challenge in identifying the Out-Of-The-Loop (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 [16, 20].

More specifically, the use of linear temporal logic on finite traces (LTL_f) formulas is proposed to characterize behaviours from observed finite traces [42]. 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 [43, 44]. These execution logs, composed of traces that comprise a series of system states[45], are interpreted over finite traces using LTL_f formulas [46]. 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.

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 [16]. A Bayesian probabilistic model is also employed to infer contrastive explanations that delineate differences between various traces for learning LTL_f formulas [19]. 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 [47]. However, these approaches may pose challenges in the analysis of dynamic and complex eye-tracking data as they either presuppose a noise-free environment [48, 16], limit the hypothesis space by LTL_f template [20, 49], or are subject to the high complexity inherent in MaxSAT [47].

This paper proposes the integration of Graph Neural Network (GNN) inference to support the inference of LTL_f, building on the work of Luo et al. [46]. 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.

2.3. Parse attention distribution through eye-tracking

Dynamic gaze movements are 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 [50, 51]. 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 [52, 53, 54]. For example, Schnebelen et al. partitioned the driving scene into 13 AOIs based on the positions and functionalities in road traffic [18]. 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 [55]. 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 [56]. 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 [18]. 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 [57]. 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.

Nevertheless, 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 [58]. 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 [59], while it relies on the complete observation of all aircrafts (AOIs) on the display. Other studies encode the dwells into vectors of fixated length [60] or calculate the transition matrices across the AOIs [61, 56], 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.

To address the gap, this paper proposes a method for encoding the visual attention of the pilots and presents a framework called *Visual Atten-*

tion LTL_f for Identifying OOTL (VALIO), which aims to achieve explainable OOTL identification using the eye-tracking data from pilots' visual behaviours.

3. Visual Attention LTL_f for Identifying OOTL (VALIO)

The *Visual Attention LTL_f Identifying OOTL (VALIO)* framework, as illustrated in Figure 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 this section.

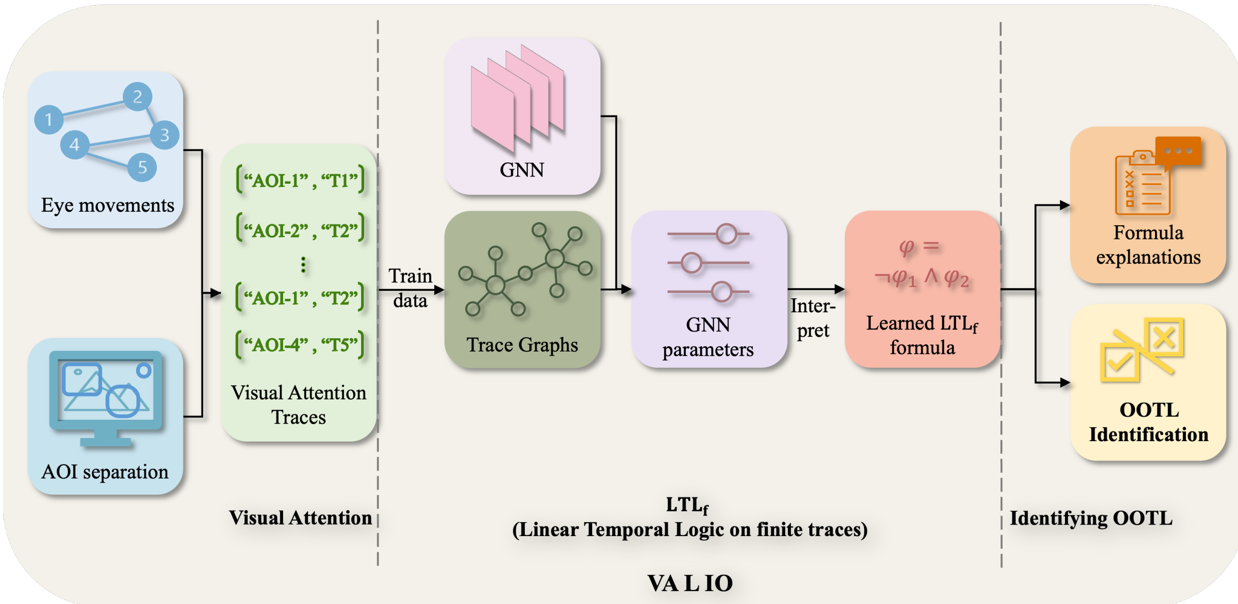


Figure 1: Framework of Visual Attention LTL_f for Identifying OOTL (VALIO)

3.1. Visual Attention Traces

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 [62].

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. 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 [50, 63]. 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 (1) and (2):

$$a_i \in \{AOI_1, AOI_2, AOI_3, ...\} \tag{1}$$

$$t_i \in \{T_1, T_2, ..., T_\tau\} \tag{2}$$

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 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 (3) below:

$$d_{-1} = \begin{cases} d_{n-1}, & |\ |d_0, ..., d_{n-1}| - 20sec\ | < \ |\ |d_0, ..., d_n| - 20sec\ | \\ d_n, & |\ |d_0, ..., d_{n-1}| - 20sec\ | > \ |\ |d_0, ..., d_n| - 20sec\ | \end{cases} \tag{3}$$

where $|d_0, ..., d_n|$ represents the total duration from the first dwell (d_0) to the n^{th} 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.

3.2. 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 2, the VAT in the previous steps can be regarded as a special form of LTL_f traces with the following characteristics:

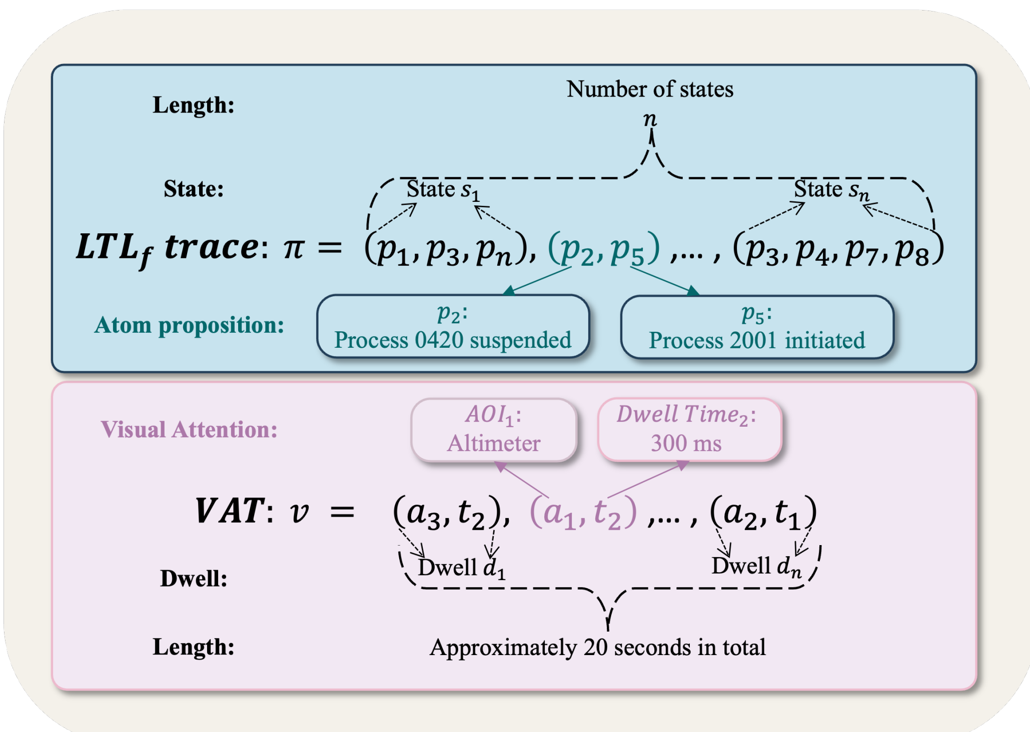


Figure 2: Composition of LTL_f traces and VAT

- **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. [46]. 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.

3.3. 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. Case study

This study 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.

4.1. Experiment design and data collection

The experimental framework utilized a Cessna 172 simulator, coupled with Microsoft Flight Simulator [64] to create a realistic aerial environment.

A Tobii Pro Glasses 3 [65], 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 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.

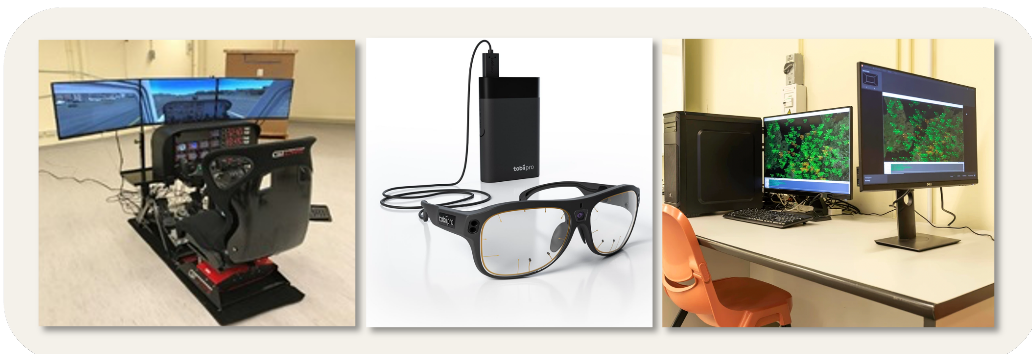


Figure 3: Experiment apparatus

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) [66]: they only need to conduct a Straight-and-Level flight [67] 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 practicing 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 [68, 18, 69]. 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 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 [70]. 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.

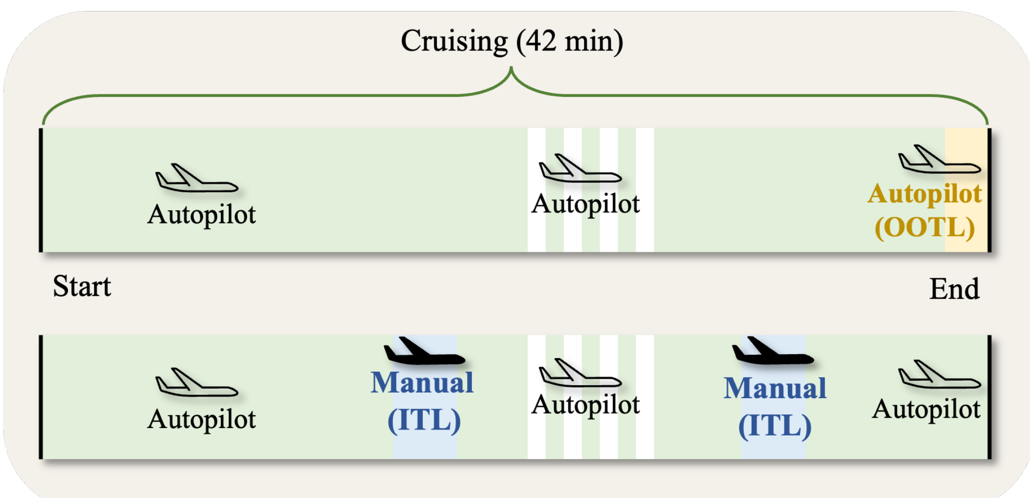


Figure 4: Flight tasks

4.2. 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 [71, 72, 73]. We tested four methods for 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.

4.2.1. Areas of Interests in the cockpit

The Cessna simulator provides instruments on the panel and the simulated view of the flight. Figure 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 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.

4.2.2. 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 [74]. 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.

Figure 5: Screenshot of Tobii Pro Lab

- *Logarithm(Ln)*: Define the bin edges using logarithm to the base of the mathematical constant e. Considering $e^{4.6} \approx 100$ms (T_1 level), the rest bin edges are $e^{5.6}$ (≈ 270ms), $e^{6.6}$ (≈ 735ms), and $e^{7.6}$ (≈ 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 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)

4.2.3. Visual Attention Traces

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 in Figure 7. Based on the previously established labels,

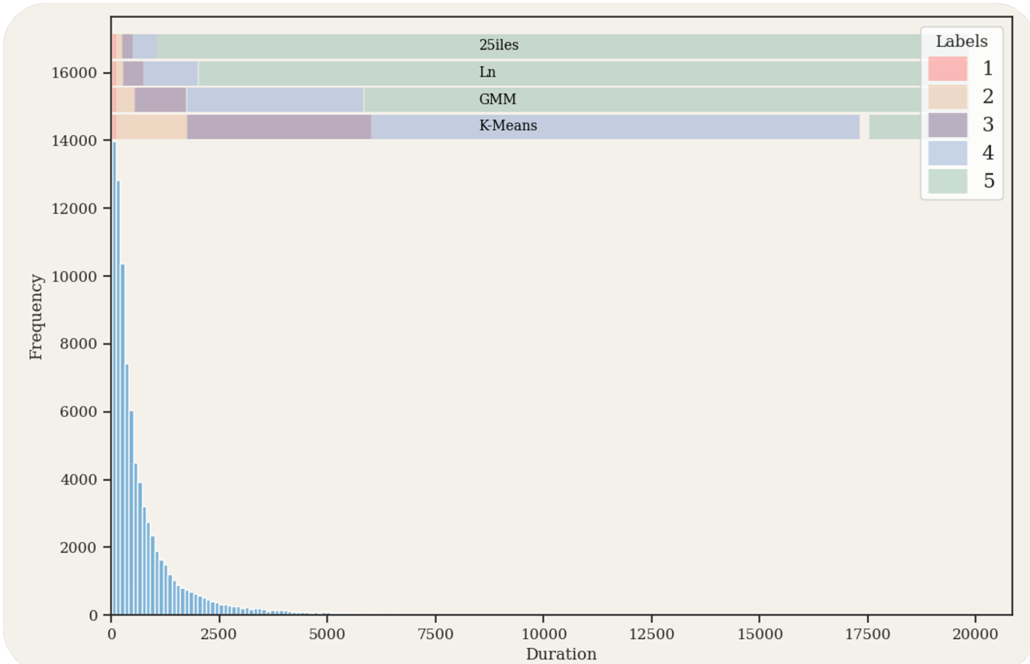


Figure 6: Division of dwell times using different methods

the VATs were categorized into OOTL status and ITL status. Based on the work of Luo et al. [46], 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. 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. Precision, 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.

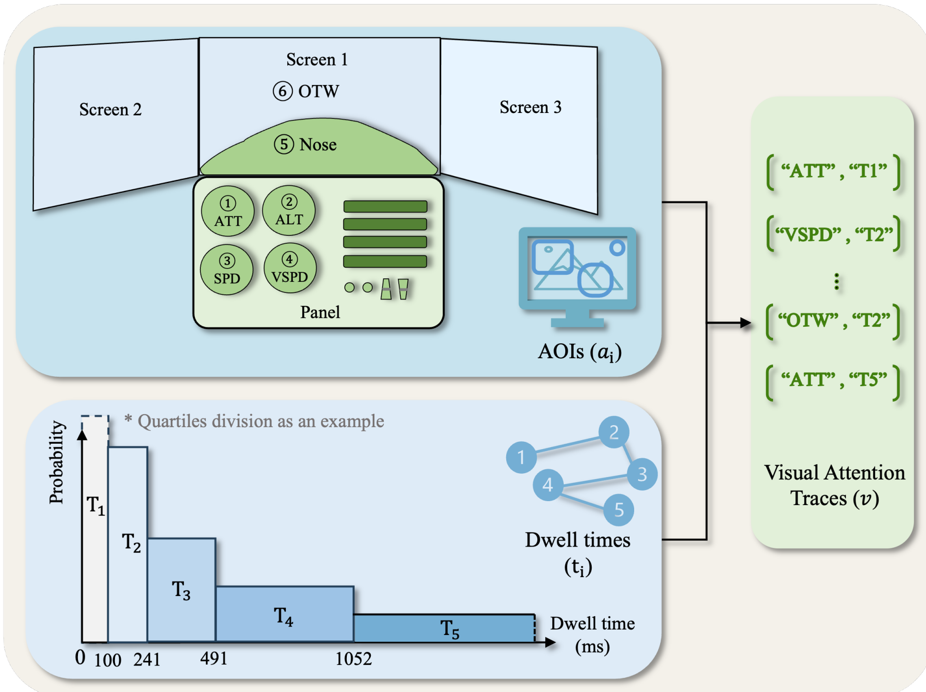


Figure 7: Temporal-spatial information in the VATs

4.3.1. 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 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.

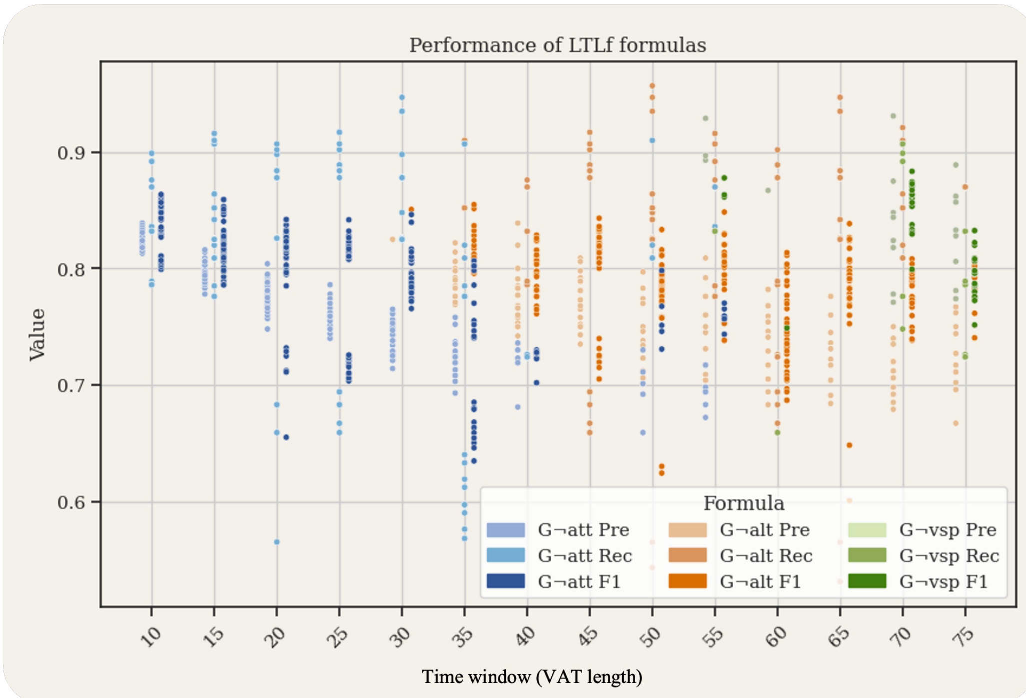


Figure 8: Distribution and performance of LTL_f formulas

- "*G¬ 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¬ 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¬ 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.

4.3.2. 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 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.

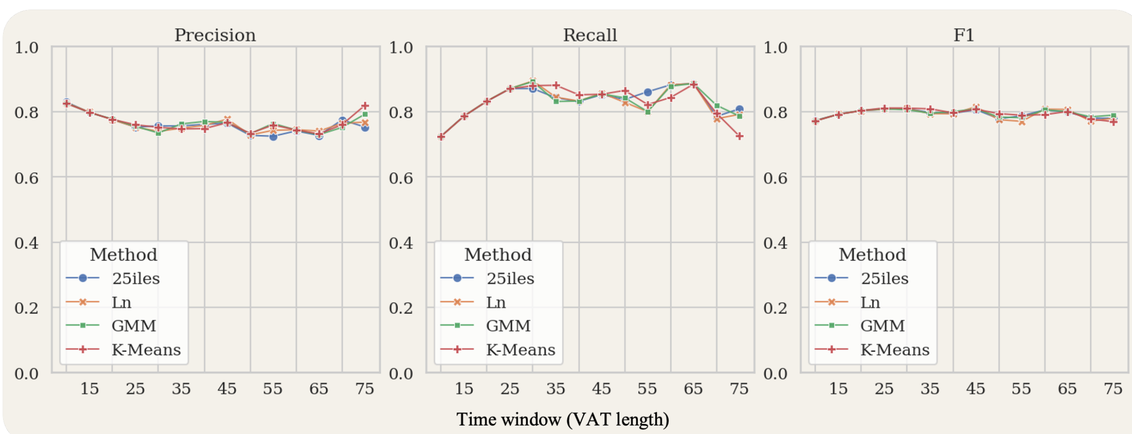


Figure 9: VALIO performance using different dwell time division methods

4.3.3. 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 Perceptron (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 1.

The performance of the compared methods is shown in Figure 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.

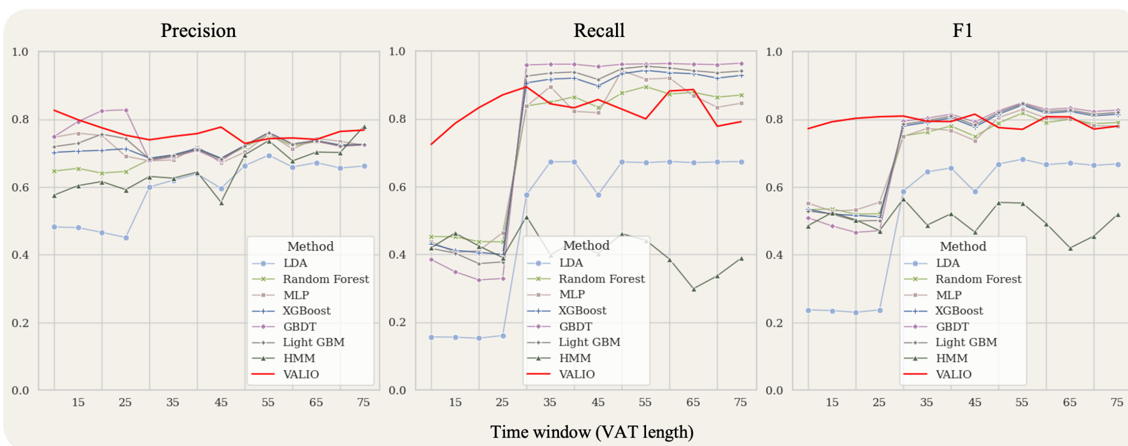


Figure 10: Performance of VALIO compared with other methods

4.4. 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.

Table 1: Machine learning models for comparison and the major hyperparameters

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

Meanwhile, the comparison 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.

4.4.1. VALIO with different VAT lengths and different 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 [75, 76].

These three formulas were recognized with the performance decline of the previous formula as shown in Figure 8, resulting in a stable overall performance of the VALIO methods across time windows.

4.4.2. VALIO and other classification methods

As shown in Figure 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 10 when the time window was longer than 30 seconds. Another observation from Figure 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.

4.4.3. Summary

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.

5. Conclusion

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. 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.

Appendix A. Abbreviations

Acknowledgements

This work was supported by the Hong Kong Polytechnic University under Grant P0038827 and Grant P0038933. This study has been granted human ethics approval from the PolyU Institutional Review Board of The Hong Kong Polytechnic University (IRB Reference Number: HSEARS20211117002).

AOI	Are of Interest
ASRS	Aviation Safety Reporting Systems
EEG	Electroencephalogram
fNIRS	functional Near-Infrared Spectroscopy
FMS	Flight Management Systems
GBDT	Gradient Boosting Decision Tree
GNN	Graph Neural Network
IDA	Information-Decision-Action (cognitive model)
LDA	Linear Discriminant Analysis
LightGBM	Light Gradient Boosting Machine
LTL	Linear Temporal Logic
LTL_f	Linear Temporal Logic on finite traces
MaxSAT	Maximum Satisfiability problem
MLP	Multilayer Perceptron
NASA	National Aeronautics and Space Administration
RF	Random Forest
SA	Situation Awareness
OOTL	Out-Of-The-Loop
VALIO	Visual Attention LTL_f for Identifying OOTL
VAT	Visual Attention Trace
XGBoost	eXtreme Gradient Boosting

Data Availability Statement

Data will be made available on request.

References

[1] N. R. Council, Taking Flight: Education and Training for Aviation Careers, The National Academies Press, Washington, DC, 1997. URL: https://nap.nationalacademies.org/catalog/5433/taking-flight-education-and-training-for-aviation-careers. doi:10.17226/5433.

[2] R. P. G. Collinson, Autopilots and flight management systems, in: R. P. G. Collinson (Ed.), Introduction to Avionics, Springer Netherlands, Dordrecht, 1996, pp. 366–405. URL: http://link.springer.com/10.1007/978-94-011-0007-6_8. doi:10.1007/978-94-011-0007-6_8.

[3] P. Arico, G. Borghini, G. Di Flumeri, S. Bonelli, A. Golfetti, I. Graziani, S. Pozzi, J.-P. Imbert, G. Granger, R. Benhacene, D. Schaefer, F. Babiloni, Human Factors and Neurophysiological Metrics in Air Traffic Control: A Critical Review, IEEE Reviews in Biomedical Engineering 10 (2017) 250–263. URL: http://ieeexplore.ieee.org/document/7898410/. doi:10.1109/RBME.2017.2694142.

[4] M. R. Endsley, D. B. Kaber, Level of automation effects on performance, situation awareness and workload in a dynamic control task, Ergonomics 42 (1999) 462–492. URL: https://www.tandfonline.com/doi/full/10.1080/001401399185595. doi:10.1080/001401399185595.

[5] F. Li, C.-H. Chen, C.-H. Lee, S. Feng, Artificial intelligence-enabled non-intrusive vigilance assessment approach to reducing traffic controller's human errors, Knowledge-Based Systems (2021) 108047. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950705121011369. doi:10.1016/j.knosys.2021.108047.

[6] M. R. Endsley, E. O. Kiris, The Out-of-the-Loop Performance Problem and Level of Control in Automation, Human Factors: The Journal of the Human Factors and Ergonomics Society 37 (1995) 381–394. URL: http://journals.sagepub.com/doi/10.1518/001872095779064555. doi:10.1518/001872095779064555.

[7] K. Gerbert, R. Kemmler, The causes of causes: determinants and background variables of human factor incidents and accidents, Ergonomics 29 (1986) 1439–1453. URL: http://www.tandfonline.com/doi/abs/10.1080/00140138608967257. doi:10.1080/00140138608967257.

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

[9] J. Gouraud, A. Delorme, B. Berberian, Autopilot, Mind Wandering, and the Out of the Loop Performance Problem, Frontiers in Neuroscience 11 (2017) 541. URL: http://journal.frontiersin.org/article/10.3389/fnins.2017.00541/full. doi:10.3389/fnins.2017.00541.

[10] N. Merat, B. Seppelt, T. Louw, J. Engström, J. D. Lee, E. Johansson, C. A. Green, S. Katazaki, C. Monk, M. Itoh, D. McGehee, T. Sunda, K. Unoura, T. Victor, A. Schieben, A. Keinath, The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications, Cognition, Technology & Work 21 (2019) 87–98. URL: http://link.springer.com/10.1007/s10111-018-0525-8. doi:10.1007/s10111-018-0525-8.

[11] L. Mengtao, L. Fan, X. Gangyan, H. Su, Leveraging eye-tracking technologies to promote aviation safety- A review of key aspects, challenges, and future perspectives, Safety Science 168 (2023) 106295. URL: https://www.sciencedirect.com/science/article/pii/S0925753523002370. doi:https://doi.org/10.1016/j.ssci.2023.106295.

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

[13] P. Cui, S. Athey, Stable learning establishes some common ground between causal inference and machine learning, Nature Machine Intelligence 4 (2022) 110–115. URL: https://www.nature.com/articles/s42256-022-00445-z. doi:10.1038/s42256-022-00445-z.

[14] W. Saeed, C. Omlin, Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities, Knowledge-Based Systems 263 (2023) 110273. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950705123000230. doi:10.1016/j.knosys.2023.110273.

[15] F. Li, C.-H. Chen, C.-H. Lee, L.-P. Khoo, A User Requirement-driven Approach Incorporating TRIZ and QFD for Designing a Smart Vessel Alarm System to Reduce Alarm Fatigue, Journal of Navigation 73 (2020) 212–232. URL: https://www.cambridge.org/core/product/identifier/S0373463319000547/type/journal_article. doi:10.1017/S0373463319000547.

[16] A. Camacho, S. A. McIlraith, Learning Interpretable Models Expressed in Linear Temporal Logic, Proceedings of the International Conference on Automated Planning and Scheduling 29 (2021) 621–630. URL: https://ojs.aaai.org/index.php/ICAPS/article/view/3529. doi:10.1609/icaps.v29i1.3529.

[17] C. Conati, C. Merten, Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation, Knowledge-Based Systems 20 (2007) 557–574. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950705107000469. doi:10.1016/j.knosys.2007.04.010.

[18] D. Schnebelen, C. Charron, F. Mars, Estimating the out-of-the-loop phenomenon from visual strategies during highly automated driving, Accident Analysis & Prevention 148 (2020) 105776. URL: https://linkinghub.elsevier.com/retrieve/pii/S0001457520315967. doi:10.1016/j.aap.2020.105776.

[19] J. Kim, C. Muise, A. Shah, S. Agarwal, J. Shah, Bayesian Inference of Linear Temporal Logic Specifications for Contrastive Explanations, in: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence Organization, Macao, China, 2019, pp. 5591–5598. URL: https://www.ijcai.org/proceedings/2019/776. doi:10.24963/ijcai.2019/776.

[20] C. Lemieux, D. Park, I. Beschastnikh, General LTL Specification Mining (T), in: 2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE), IEEE, Lincoln, NE, USA, 2015, pp. 81–92. URL: http://ieeexplore.ieee.org/document/7371998/. doi:10.1109/ASE.2015.71.

[21] X. Zheng, C. Yu, M. Zhang, Lifelong reinforcement learning with temporal logic formulas and reward machines, Knowledge-Based Systems

257 (2022) 109650. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950705122008358. doi:10.1016/j.knosys.2022.109650.

[22] T. J. Etherington, L. J. Kramer, K. H. Shish, S. D. Young, E. T. Evans, T. S. Daniels, Increasing Pilot's Understanding of Future Automation State - An Evaluation of an Automation State and Trajectory Prediction System, in: 2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC), IEEE-AIAA Digital Avionics Systems Conference, IEEE; AIAA; AIAA Digital Avion Tech Comm; IEEE Aerosp & Elect Syst Soc; NASA; Boeing; Honeywell; Presagis; MITRE Corp; Mark III Syst; Green Hills Software; Wind River; AFuzion; Gen Atom Aeronaut Syst Inc; Reliable Robot, 2019. ISSN: 2155-7195.

[23] D. B. Kaber, M. R. Endsley, Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety, Process Safety Progress 16 (1997) 126–131. URL: https://onlinelibrary.wiley.com/doi/10.1002/prs.680160304. doi:10.1002/prs.680160304.

[24] G. Di Flumeri, F. De Crescenzio, B. Berberian, O. Ohneiser, J. Kramer, P. Aricò, G. Borghini, F. Babiloni, S. Bagassi, S. Piastra, Brain–Computer Interface-Based Adaptive Automation to Prevent Out-Of-The-Loop Phenomenon in Air Traffic Controllers Dealing With Highly Automated Systems, Frontiers in Human Neuroscience 13 (2019) 296. URL: https://www.frontiersin.org/article/10.3389/fnhum.2019.00296/full. doi:10.3389/fnhum.2019.00296.

[25] B. Lorenz, F. Di Nocera, S. Röttger, R. Parasuraman, The Effects of Level of Automation on the Out-of-the-Loop Unfamiliarity in a Complex Dynamic Fault-Management Task during Simulated Spaceflight Operations, Proceedings of the Human Factors and Ergonomics Society Annual Meeting 45 (2001) 44–48. URL: http://journals.sagepub.com/doi/10.1177/154193120104500209. doi:10.1177/154193120104500209.

[26] T. Louw, N. Merat, Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation, Transportation Research Part C: Emerging Technologies 76 (2017) 35–50. URL: https://linkinghub.elsevier.com/retrieve/pii/S0968090X17300013. doi:10.1016/j.trc.2017.01.001.

[27] A. Sebok, C. D. Wickens, Implementing Lumberjacks and Black Swans Into Model-Based Tools to Support Human–Automation Interaction, Human Factors: The Journal of the Human Factors and Ergonomics Society 59 (2017) 189–203. URL: http://journals.sagepub.com/doi/10.1177/0018720816665201. doi:10.1177/0018720816665201.

[28] C. D. Wickens, Cognitive factors in aviation., in: Handbook of applied cognition., John Wiley & Sons Ltd, New York, NY, US, 1999, pp. 247–282.

[29] N. B. Sarter, G. Meyer, How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control, Human Factors: The Journal of the Human Factors and Ergonomics Society 37 (1995) Road Vehicle Automation 2. URL: 978-3-319-19077-8978-3-319-19078-5. doi:10.1518/001872095779049516.

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

[31] D. A. Norman, The Human Side of Automation, in: G. Meyer, S. Beiker (Eds.), Road Vehicle Automation 2, Springer International Publishing, Cham, 2015, pp. 73–79. URL: https://link.springer.com/10.1007/978-3-319-19078-5_7. doi:10.1007/978-3-319-19078-5_7, series Title: Lecture Notes in Mobility.

[32] B. Berberian, B. Somon, A. Sahaï, J. Gouraud, The out-of-the-loop Brain: A neuroergonomic approach of the human automation interaction, Annual Reviews in Control 44 (2017) 303–315. URL: https://linkinghub.elsevier.com/retrieve/pii/S1367578817301116. doi:10.1016/j.arcontrol.2017.09.010.

[33] F. Li, C.-H. Lee, C.-H. Chen, L. P. Khoo, Hybrid data-driven vigilance model in traffic control center using eye-tracking data and context data, Advanced Engineering Informatics 42 (2019) 100940. URL: https://linkinghub.elsevier.com/retrieve/pii/S1474034619300540. doi:10.1016/j.aei.2019.100940.

[34] A. Hamann, N. Carstengerdes, Assessing the development of mental fatigue during simulated flights with concurrent EEG-fNIRS measure-

ment, Scientific Reports 13 (2023) 4738. URL: https://www.nature.com/articles/s41598-023-31264-w. doi:10.1038/s41598-023-31264-w.

[35] M. Mittner, W. Boekel, A. M. Tucker, B. M. Turner, A. Heathcote, B. U. Forstmann, When the Brain Takes a Break: A Model-Based Analysis of Mind Wandering, The Journal of Neuroscience 34 (2014) 16286–16295. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.2062-14.2014. doi:10.1523/JNEUROSCI.2062-14.2014.

[36] R. Bixler, S. D'Mello, Automatic gaze-based user-independent detection of mind wandering during computerized reading, User Modeling and User-Adapted Interaction 26 (2016) 33–68. URL: http://link.springer.com/10.1007/s11257-015-9167-1. doi:10.1007/s11257-015-9167-1.

[37] F. Li, C.-H. Chen, P. Zheng, S. Feng, G. Xu, L. P. Khoo, An explorative context-aware machine learning approach to reducing human fatigue risk of traffic control operators, Safety Science 125 (2020) 104655. URL: https://linkinghub.elsevier.com/retrieve/pii/S0925753520300527. doi:10.1016/j.ssci.2020.104655.

[38] Z. Li, R. Li, L. Yuan, J. Cui, F. Li, A benchmarking framework for eye-tracking-based vigilance prediction of vessel traffic controllers, Engineering Applications of Artificial Intelligence 129 (2024) 107660. URL: https://linkinghub.elsevier.com/retrieve/pii/S0952197623018444. doi:10.1016/j.engappai.2023.107660.

[39] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, Explainable AI: A Review of Machine Learning Interpretability Methods, Entropy 23 (2020) 18. URL: https://www.mdpi.com/1099-4300/23/1/18. doi:10.3390/e23010018.

[40] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444. URL: https://www.nature.com/articles/nature14539. doi:10.1038/nature14539.

[41] R. Polikar, Ensemble Learning, in: C. Zhang, Y. Ma (Eds.), Ensemble Machine Learning, Springer New York, New York, NY, 2012, pp. 1–34. URL: https://link.springer.com/10.1007/978-1-4419-9326-7_1. doi:10.1007/978-1-4419-9326-7_1.

[42] D. Kasenberg, M. Scheutz, Interpretable apprenticeship learning with temporal logic specifications, in: 2017 IEEE 56th Annual Conference on Decision and Control (CDC), IEEE, Melbourne, Australia, 2017, pp. 4914–4921. URL: http://ieeexplore.ieee.org/document/8264386/. doi:10.1109/CDC.2017.8264386.

[43] M. Gabel, Z. Su, 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, ACM, Atlanta Georgia, 2008, pp. 339–349. URL: https://dl.acm.org/doi/10.1145/1453101.1453150. doi:10.1145/1453101.1453150.

[44] D. Lo, S. Khoo, C. Liu, Mining temporal rules for software maintenance, Journal of Software Maintenance and Evolution: Research and Practice 20 (2008) 227–247. URL: https://onlinelibrary.wiley.com/doi/10.1002/smr.375. doi:10.1002/smr.375.

[45] J. Shi, J. Xiong, Y. Huang, General past-time linear temporal logic specification mining, CCF Transactions on High Performance Computing 3 (2021) 393–406. URL: https://link.springer.com/10.1007/s42514-021-00079-4. doi:10.1007/s42514-021-00079-4.

[46] W. Luo, P. Liang, J. Du, H. Wan, B. Peng, D. Zhang, Bridging LTLf Inference to GNN Inference for Learning LTLf Formulae, Proceedings of the AAAI Conference on Artificial Intelligence 36 (2022) 9849–9857. URL: https://ojs.aaai.org/index.php/AAAI/article/view/21221. doi:10.1609/aaai.v36i9.21221.

[47] J.-R. Gaglione, D. Neider, R. Roy, U. Topcu, Z. Xu, Learning Linear Temporal Properties from Noisy Data: A MaxSAT-Based Approach, in: Z. Hou, V. Ganesh (Eds.), Automated Technology for Verification and Analysis, volume 12971, Springer International Publishing, Cham, 2021, pp. 74–90. URL: https://link.springer.com/10.1007/978-3-030-88885-5_6. doi:10.1007/978-3-030-88885-5_6, series Title: Lecture Notes in Computer Science.

[48] D. Neider, I. Gavran, Learning Linear Temporal Properties, in: 2018 Formal Methods in Computer Aided Design (FMCAD), IEEE, Austin, TX, 2018, pp. 1–10. URL: https://ieeexplore.ieee.org/document/8603016/. doi:10.23919/FMCAD.2018.8603016.

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

[50] K. Holmqvist, R. Andersson, Eye-tracking: A comprehensive guide to methods, paradigms and measures, Oxford University Press, 2011. URL: http://ukcatalogue.oup.com/product/9780199697083.do.

[51] R. S. Hessels, C. Kemner, C. Van Den Boomen, I. T. C. Hooge, The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli, Behavior Research Methods 48 (2016) 1694–1712. URL: http://link.springer.com/10.3758/s13428-015-0676-y. doi:10.3758/s13428-015-0676-y.

[52] B. Mahanama, Y. Jayawardana, S. Rengarajan, G. Jayawardena, L. Chukoskie, J. Snider, S. Jayarathna, Eye Movement and Pupil Measures: A Review, Frontiers in Computer Science 3 (2022) 733531. URL: https://www.frontiersin.org/articles/10.3389/fcomp.2021.733531/full. doi:10.3389/fcomp.2021.733531.

[53] F. Li, C.-H. Lee, S. Feng, A. Trappey, F. Gilani, Prospective on Eye-Tracking-based Studies in Immersive Virtual Reality, in: 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), IEEE, Dalian, China, 2021, pp. 861–866. URL: https://ieeexplore.ieee.org/document/9437692/. doi:10.1109/CSCWD49262.2021.9437692.

[54] Z. Li, Z. Li, F. Li, Visual Attention Analytics for Individual Perception Differences and Task Load-Induced Inattentional Blindness, in: P.-L. P. Rau (Ed.), Cross-Cultural Design, volume 14023, Springer Nature Switzerland, Cham, 2023, pp. 71–83. URL: https://link.springer.com/10.1007/978-3-031-35939-2_6. doi:10.1007/978-3-031-35939-2_6, series Title: Lecture Notes in Computer Science.

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

volume 8532 of *Lecture Notes in Computer Science*, 2014, pp. 325–334. ISSN: 0302-9743.

[56] A. Haslbeck, B. Zhang, I spy with my little eye: Analysis of airline pilots' gaze patterns in a manual instrument flight scenario, Applied Ergonomics 63 (2017) 62–71. URL: https://linkinghub.elsevier.com/retrieve/pii/S000368701730073X. doi:10.1016/j.apergo.2017.03.015.

[57] G. Ziv, Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation, The International Journal of Aviation Psychology 26 (2017) 75–104. doi:10.1080/10508414.2017.1313096.

[58] G. Jayawardena, S. Jayarathna, Automated Filtering of Eye Movements Using Dynamic AOI in Multiple Granularity Levels:, International Journal of Multimedia Data Engineering and Management 12 (2021) 49–64. URL: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJMDEM.2021010104. doi:10.4018/IJMDEM.2021010104.

[59] S. N. McClung, Z. Kang, Characterization of visual scanning patterns in air traffic control, Computational intelligence and neuroscience 2016 (2016).

[60] V. Peysakhovich, W. Ledegang, M. Houben, E. Groen, Classification of flight phases based on pilots' visual scanning strategies, in: 2022 Symposium on Eye Tracking Research and Applications, 2022, pp. 1–7.

[61] Z. Kang, S. J. Landry, An eye movement analysis algorithm for a multi-element target tracking task: Maximum transition-based agglomerative hierarchical clustering, IEEE Transactions on Human-Machine Systems 45 (2014) 13–24.

[62] F. Li, S. Katsumata, C.-H. Lee, Q. Ye, W. D. Dahana, R. Tu, X. Li, Autoencoder-Enabled Potential Buyer Identification and Purchase Intention Model of Vacation Homes, IEEE Access 8 (2020) 212383–212395. URL: https://ieeexplore.ieee.org/document/9258913/. doi:10.1109/ACCESS.2020.3037920.

[63] A. Mohanty, T. J. Sussman, Top-down modulation of attention by emotion, Frontiers in Human Neuroscience 7 (2013). URL: http://journal.frontiersin.org/article/10.3389/fnhum.2013.00102/abstract. doi:10.3389/fnhum.2013.00102.

[64] Microsoft, Microsoft Flight Simulator, 2024. URL: https://www.flightsimulator.com/.

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

[66] Q. Li, K. K. Ng, C. Y. Yiu, X. Yuan, C. K. So, C. C. Ho, Securing air transportation safety through identifying pilot's risky vfr flying behaviours: An eeg-based neurophysiological modelling using machine learning algorithms, Reliability Engineering & System Safety 238 (2023) 109449.

[67] F. A. Administration, Airplane flying handbook (FAA-H-8083-3A), Skyhorse Publishing Inc., 2011.

[68] A. Feldhütter, C. Gold, S. Schneider, K. Bengler, How the Duration of Automated Driving Influences Take-Over Performance and Gaze Behavior, in: C. M. Schlick, S. Duckwitz, F. Flemisch, M. Frenz, S. Kuz, A. Mertens, S. Mütze-Niewöhner (Eds.), Advances in Ergonomic Design of Systems, Products and Processes, Springer Berlin Heidelberg, Berlin, Heidelberg, 2017, pp. 309–318. URL: http://link.springer.com/10.1007/978-3-662-53305-5_22. doi:10.1007/978-3-662-53305-5_22.

[69] M. Lyu, F. Li, C.-H. Lee, The Effects of Adaptive Automation on Pilots' Flight Control Performance and Visual Attention Distribution, in: P. Koomsap, A. Cooper, J. Stjepandić (Eds.), Advances in Transdisciplinary Engineering, IOS Press, 2023. URL: https://ebooks.iospress.nl/doi/10.3233/ATDE230600. doi:10.3233/ATDE230600.

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

[71] P. Gontar, H. Homans, M. Rostalski, J. Behrend, F. Dehais, K. Bengler, Are pilots prepared for a cyber-attack? a human factors approach to the experimental evaluation of pilots' behavior, Journal of Air Transport Management 69 (2018) 26–37.

[72] J. Kee, M. Knuth, J. N. Lahey, M. A. Palma, Does eye-tracking have an effect on economic behavior?, Plos one 16 (2021) e0254867.

[73] J. Aust, A. Mitrovic, D. Pons, Assessment of the effect of cleanliness on the visual inspection of aircraft engine blades: an eye tracking study, Sensors 21 (2021) 6135.

[74] F. Li, C.-H. Chen, G. Xu, L. P. Khoo, Y. Liu, Proactive mental fatigue detection of traffic control operators using bagged trees and gaze-bin analysis, Advanced Engineering Informatics 42 (2019) 100987. URL: https://linkinghub.elsevier.com/retrieve/pii/S1474034619305609. doi:10.1016/j.aei.2019.100987.

[75] L. L. Di Stasi, M. B. McCamy, S. Martinez-Conde, E. Gayles, C. Hoare, M. Foster, A. Catena, S. L. Macknik, Effects of long and short simulated flights on the saccadic eye movement velocity of aviators, Physiology & Behavior 153 (2016) 91–96. doi:10.1016/j.physbeh.2015.10.024.

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